Learning Stroke Treatment Progression Models for an MDP Clinical Decision Support System [*]

Dan C. Coroian[†] Kris Hauser[‡]

Abstract

This paper describes a clinical decision support framework in multi-step health care domains that can dynamically recommend optimal treatment plans with respect to both patient outcomes and expected treatment cost. Our system uses a modified POMDP framework in which hidden states are not explicitly modeled, but rather, probabilistic models for predicting future observables given observation and action histories are learned directly from electronic health record (EHR) data. High quality treatment recommendations are found using a sampling-based tree growing approach which produces good results despite only exploring a fraction of the observation and action spaces. We describe the application of the approach to an ischemic stroke domain with clinical trial data (International Stroke Trial Dataset, 1993-1996). The dataset is of moderate size (N=19,435) and exhibits many characteristics of real EHR data, including noise, missing values, and idiosyncratic coding. The system's predictive model was chosen using cross-validated model selection from a set of several candidate learning methods, including logistic regression, Naïve Bayes, Bayes nets, and random forests. Simulations suggest that the optimized decisions improve patient outcomes, such as 6-month survival rate, compared to the decisions of human doctors during the study.

Keywords: Health care, decision-making, Markov Decision Processes, time-series models, optimization, machine learning.

1 Introduction

The United States health care system is by far the most expensive in the world while delivering worse patient outcomes than other industrialized, Western nations [7]. One approach to this problem is a "big data" approach in which algorithms mine the vast amounts of electronic health record (EHR), clinical trial, and genome data to make personalized healthcare recommendations to both clinicians and patients. This approach has the potential to improve both quality and cost of care [3, 10, 11].

This paper presents an entirely data-driven clinical decision support system (CDSS) for the treatment of ischemic stroke. Stroke is the fourth-highest cause of death and the leading cause of disability in the U.S. [18], and even minor improvements in stroke treatment can lead to major economic savings because of the long-term care typically required. Disagreements about best practices may take years to resolve via clinical studies and then to disseminate to clinical practice, particularly outside of state-of-the-art stroke centers, and a CDSS has the potential to integrate vast amounts of real-time data to mine for and disseminate best practices. However, stroke is a challenging domain for clinical decision support systems because of the large number of clinical variables, medications and dosages that may be administered, and surgical options. It also requires both acute treatment upon initial stroke as well as long-term management due to the high rate of recurrence [15].

Our system is based on clinical data from the International Stroke Trial (IST) study [16, 17]. This large study was conducted between 1991-1996 on 19,435 participants across 40 countries to assess the benefits of aspirin and/or heparin therapy in patients who have suffered an ischemic stroke. For each patient, data were collected at three distinct temporal stages, yielding over 100 clinically-relevant variables. Our CDSS learns treatment progression models which are then used to optimize treatment decisions at each stage. These models are temporal, relying on observations and treatments from previous stages in order to make predictions about the future, as well as probabilistic, allowing the CDSS to deliberate about tradeoffs between uncertain outcomes.

The key contribution of this paper is a purely data-driven method for learning and applying treatment progression models in a partially-observable decision-theoretic framework. From an EHR dataset we estimate conditional probabilities of future observables, such as indicators of disease progression or patient outcomes, given treatment history. Our approach has many of the strengths of the Partially Observable Markov Decision Process (POMDP) [1] framework in being able to optimize multi-step temporal treatments with hidden

[*] Supported by NSF Smart and Connected Health Grant #1343940.

[†] School of Informatics and Computing, Indiana University, 901 E. 10th St, Bloomington, IN 47408, USA.

[‡] Pratt School of Engineering, Duke University, Gross Hall Rm 379, 140 Science Dr, Durham, NC 27708, USA.

variables, but avoids the major difficulties of defining the temporal dynamics of unobserved state variables.

The temporal modeling and decision problems faced in the stroke domain are still very large and intractable for traditional techniques. To apply our approach successfully in practice, we introduce two technical contributions. First, the number of possible treatment policies is enormous, and so learning faces a combinatorial explosion over dozens of treatment and observation variables. Overfitting is a major concern. We address this by performing cross-validated model selection to find the best predictive model from a number of machine learning techniques. Second, the MDP problem is enormous and is intractable for exact solution techniques. To handle the large observation space, we apply approximate Monte-Carlo sampling techniques. To handle the huge number of possible combinations of actions, we mine the dataset for clinically-relevant action combinations to avoid recommending nonsensical treatments that do not exist in clinical usage.

The CDSS is applied to optimize treatments that penalize costs and negative patient outcomes, like death or long-term dependence. Simulations suggest it may improve stroke survival rate from 80% to 84% and improve the likelihood of full recovery from 19% to 30%.

2 Background and Related Work

Previous work identified four key features of a CDSS which were significantly correlated with improving clinical practice, namely, using a computerized CDSS, providing support automatically as part of the clinician's pre-existing workflow, doing so at the time and place of decision-making, and offering alternative suggestions instead of simple assessments [11]. The system proposed here is consistent with all of these desirable features. In another similar analysis, other key factors include speed of support, doctor usability, and making recommendations which are as similar as possible to the traditional course of treatment to increase the likelihood of adoption [3]. For example, one study found that simply recommending a lower typical dose of a single commonly prescribed medication (rather than changing the treatment altogether) resulted in a $250,000 savings in just one year when implemented in a single facility [3].

Machine learning techniques have shown success with respect to both prognosis and diagnosis in a variety of medical domains. For example, Bayes nets (one of the techniques we consider) were found to be successful at modeling patient progression over multiple time steps when applied to the cardiac surgery domain [19, 20]. Another study obtained over 94% diagnosis accuracy in predicting cardiovascular disease using an ensemble method which somewhat resembles the random forest technique considered here [8]. In this paper, however, we do not commit to one model and instead perform model selection to find the best candidate based strictly on empirical evidence from the data.

AI techniques for decision-making such as MDPs and their partially-observable variants have also been applied to medical domains. They have yielded promising results in clinical decision-making tasks, including optimizing timing of living-donor liver transplantation, deciding when to medicate vs. operate in patients with Parkinson's disease, and clinical depression treatment [1, 4, 9]. POMDPs, in particular, model a large number of clinical domains well because of their capacity to handle complex decision-making tasks with stochastic relationships between hidden states, observations, and actions. Typical POMDP methods model treatment progression by assuming the existence of a hidden time-varying state variable $\mathbf{s}_t$, which changes over time according to a transition model $P(\mathbf{s}_{t+1}|\mathbf{s}_t, \mathbf{a}_t)$ and generates an observation according to an observation model $P(\mathbf{o}_t \,|\, \mathbf{s}_t)$. The POMDP then operates over belief states (probability distributions over the hidden actual state variable). However, the notion of actual "patient state" is nebulous and may not have clinical relevance. Moreover, patient state is often difficult to model since it is never directly observed in a dataset.

Instead, we *learn the treatment progression models directly from the dataset.* Specifically, we learn models for future observations based on the history of past observations and actions, in a manner similar to predictive state representations [13]. Based on the fact that state is a sufficient statistic for history, this method is no less powerful than a traditional POMDP, and is sometimes even more so because it avoids arbitrarily collapsing a rich dependence on history into a finite number of states.

However, even without explicitly modeling belief states, the large state and action space associated with the stroke domain would still be too expensive to solve precisely for the globally optimal solution. We take a Monte Carlo tree search approach which incrementally grows a probabilistic decision tree by taking samples according to the learned treatment progression models, in a manner similar to [12].

3 Methods

We present a general-purpose method for a CDSS that makes multi-step treatment recommendations directly from a dataset, using little human intervention. It consists of the following steps: 1) build *treatment progression models* from an EHR or clinical trial dataset, 2) mine the dataset for significantly used action sets, 3) define a reward function, and 4) use the models, action sets, and reward function in a Markov Decision Process that makes optimized treatment recommendations for new patients. In the stroke domain that we consider, a human analyst is only required to perform minor data cleaning, define a reward function, and choose a few parameters governing computational speed vs optimality. This section describes the treatment modeling (Sec. 3.3) and decision making techniques (Sec. 3.4) used here, as well as their application to the IST domain (Sec. 3.1).

3.1 International Stroke Trial Dataset The IST study was a randomized controlled trial aimed to determine the efficacy and safety of aspirin and/or heparin therapy in patients who had suffered an ischemic stroke. This dataset was ideal for building and testing our proposed system due to its multi-step nature, with a small number of steps. The dataset poses other challenges that make it difficult to handle using standard decision support techniques, such as differing sets of actions and observations from time step to time step, inclusion of many irrelevant variables, coding idiosyncrasies, and missing values. Furthermore, because of the trial's historical importance, we can compare the recommendations of our system to those of numerous statistical analyses and follow-up studies (see Sec. 4.3).

Patients were entered into the trial upon checking in to the hospital after a suspected ischemic stroke. At the first visit, some basic information about the patient's demographics and medical history as well as details about the stroke event were collected, and a randomized treatment condition was chosen. Before their discharge from the hospital, several other pieces of information (including final diagnosis, other treatments being undergone, and incidence of serious complications) were recorded. Finally, 6 months after their initial visit, follow-up information about their final condition and medication was recorded. The data are comprised of variables collected at three distinct visits, summarized below (see Supplement for more details):

- *First visit.* Collected at the time of entry into the trial, usually within a few hours after first onset of suspected ischemic stroke. Variables include:
    1. Demographic information, such as patient age, gender, etc.
    2. Medical information, such as systolic blood pressure, whether the patient was in atrial fibrillation, etc.
    3. Details about the stroke event, such as indicators for several deficits likely caused by the stroke.
    4. Pre-trial administration of either aspirin or heparin (within 24 hours or 3 days prior to admission into study, respectively).
    5. The initial treatment decision made; either aspirin alone, heparin alone in low or high dose conditions, both, or neither.
- *Second visit.* Collected at the time of discharge from hospital or 14 days after admission, whichever occurred first (or at the time of patient death if it occurred before this time). Variables include:
    1. Occurrence of significant medical events (including recurrence of stroke), such as pulmonary embolism, deep vein thrombosis, patient death, etc.
    2. Other treatments applied to the patient outside of the study, including steroids, haemodilution, etc. These all have a significant impact on the cost of treatment to the patient as well as the likelihood of death and recovery.
    3. Second study treatment decision; takes on the same values as the decision made in the first visit, and in most cases the value does not change from the first to the second visit.
    4. Diagnosis of initial event, which can be either ischemic stroke, hemorrhagic stroke, unknown type of stroke, or not a stroke.
- *Third visit.* Collected at either 6 months after the initial stroke event, or at patient death, whichever occurred first. Variables include:
    1. Final patient condition: 1) dead, 2) dependent (the most likely outcome), 3) not recovered, or 4) recovered. In the case of death, the cause of death is typically provided.
    2. Final patient medication; whether the patient was on either of antiplatelet or anticoagulant medication, both, or neither at the time of this visit or their death.

3.2 Mathematical Definition We define a visit or series of visits of a patient P to some health provider as a T-step *course of treatment* consisting of a sequence of clinical *observations* $\mathbf{o}_1, \ldots, \mathbf{o}_T$ which are interspersed with treatment *actions* $\mathbf{a}_1, \ldots, \mathbf{a}_{T-1}$. Each observation $\mathbf{o}_t$ may combine several individual observation variables $o_t^1, \ldots, o_t^{m_t}$ and likewise each action $\mathbf{a}_t$ may combine several individual action variables $a_t^1, \ldots, a_t^{n_t}$ (Fig. 1). Due to the nature of clinical data entry, many of these variables may be missing (e.g., 5% are missing in the IST), and due to the nature of many clinical procedures, the set of observation and action variables may vary from time step to time step.

The goal of the CDSS is to recommend, for a new patient P', high quality actions $\mathbf{a}_t$ given all of

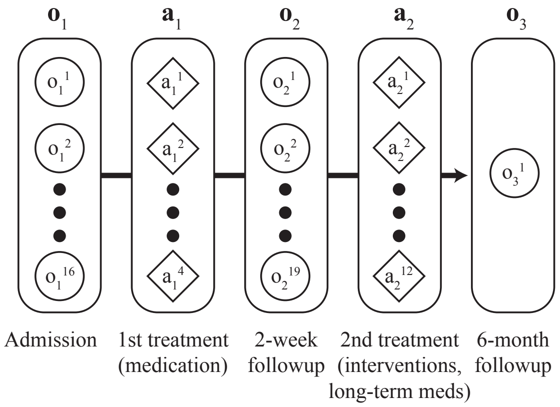


Figure 1: The temporal sequence of observations and treatments in the IST dataset consists of three observation phases and two action phases.

the patient's prior observations $\mathbf{o}_1, \ldots, \mathbf{o}_t$ and actions $\mathbf{a}_1, \ldots, \mathbf{a}_{t-1}$. This recommendation is a *policy* π that yields $\mathbf{a_t} = \pi(\mathbf{o}_1, \ldots, \mathbf{o}_t, \mathbf{a}_1, \ldots, \mathbf{a}_{t-1})$. The quality of a course of treatment x is measured by a *reward function* $R(x)$. The quality of policy π, also known as its *utility*, is measured by expected reward $E[R(x)]$ averaged over the probability space of all possible courses of treatment following π. This probability space is given by a *treatment progression distribution* that specifies a *probability distribution over* $\mathbf{o}_t$ *given the history of prior observations and actions*. In other words, if we define the *history* variables $\mathbf{h}_t = (\mathbf{o}_1, \ldots, \mathbf{o}_{t-1}, \mathbf{a}_1, \ldots, \mathbf{a}_{t-1})$, we wish to learn $P(\mathbf{o}_t \,|\, \mathbf{h}_t)$ for all $t = 1, \ldots, T$.

The dataset $\mathcal{D}$ consists of examples x_P that contain an entire course of treatment for a previously treated patient P. The purpose of our work is to 1) learn accurate approximations of the treatment progression models $\tilde{P}_t(\mathbf{o}_t \,|\, \mathbf{h}_t)$ for all $t = 1, \ldots, T$, and then 2) use the models in a POMDP-style decision-theoretic optimization of the treatment policy π.

These two steps pose several technical challenges. In the first step, the conditional density estimation problem is rather difficult. Each estimation may be conditional on dozens of variables and also predict dozens of variables: for example, the conditional distribution of the final condition depends on approx. 6×10^{18} possible distinct patient histories. Hence, data fragmentation and overfitting are issues of prime importance. Moreover, due to the nature of clinical data our models must handle training and prediction with missing values as well as heavily biased classes. In the second step, MDP solving is an additional challenge, due to the differing sets of actions and observations at each time step, as well as the huge action (e.g., $|\mathcal{A}_2| = 49,152$) and observation spaces (e.g., $|\mathcal{O}_1| = 47,239,200$) that preclude the use of exact algorithms.

3.3 Treatment Progression Models

To learn treatment progression models we compare a number of different learning methods, and perform model selection with cross-validated likelihood scoring function to choose the best. We considered the following learners:

Trivial - predicts the distribution of each variable o_t^i as unconditionally independent and simply models its distribution as its empirical distribution in $\mathcal{D}$. That is, the trivial model simply predicts the posterior probability of a class according to the proportion of examples of that class in the training set.

Bayes Network (BN) – A Bayes Network appears to be a natural fit for the clinical decision support problem due to its ability to encode conditional probabilistic dependencies among the diverse features, and the ability to encode causal dependence relations through a human's clinical expertise. We learned BNs over all variables $\mathbf{o}_1, \ldots, \mathbf{o}_T$ and $\mathbf{a}_1, \ldots, \mathbf{a}_{T-1}$, with conditional probability tables learned via standard maximum likelihood with Dirichlet priors. To define the BN structures, we compared two approaches. First, a hand-crafted BN of likely dependencies was created by the authors, with causal arcs connecting variables at one time step to variables at future time steps, and arcs were eliminated when our intuition for direct dependence was weak. We also used an automated structure learning algorithm (Tree-Augmented Naïve Bayes) that generated temporally inconsistent arcs, but performed better than the hand-coded model.

For the below learners, we model individual observations $P(o_t^i \,|\, \mathbf{h}_t)$, and assume observations are conditionally independent given history:

$$P(\mathbf{o}_t \,|\, \mathbf{h}_t) = \prod_{i=1}^{m_t} P(o_t^i \,|\, \mathbf{h}_t) \tag{3.1}$$

Naïve Bayes (NB) – a simplified BN in which only direct conditional dependencies from the target variable and each of the features are considered, with potential interactions between features ignored. The increased sparsity of the model helps generalize better/avoid overfitting more than the BN model. NB parameters were learned with a Dirichlet prior of $\alpha = 0.05$, which was found experimentally to lead to the best performance after some tuning.

Random Forest (RF) – an ensemble method that learns several weakly predictive decision trees over random subsets of the features, and a weight for each tree. To make a prediction, a weighted vote of the decision tree predictions are used to estimate posterior class probabilities. We learned an RF with 50 decision trees, each learned using the Gini impurity criterion, and each split chosen from $\sqrt{n}$ features out of n possible features. These parameters were found experimentally to give the best performance with a small amount of hand-tuning.

Logistic Regression (LR) – expresses the probability distribution as an logistic function of a linear function of the independent variables, with the parameters of the model being the linear coefficients. A multinomial variant of LR was used to model categorical variables. An assumption of LR is monotonicity of the target variable on the predictors, i.e., that the contribution of two predictor variables has a cumulative effect on the probability of the target variable, and hence is not able to model more nuanced distributions (e.g., an XOR function). The learning process numerically optimizes the linear coefficients to minimize L_2 error.

BN was implemented using the Netica package, and NB, RF, and LR were implemented using the Python library scikit-learn.

3.4 POMDP Framework We apply our models in a nontraditional POMDP approach that optimizes treatment actions without explicitly modeling states. The result is a policy $\mathbf{a}_t = \pi(\mathbf{h}_t, \mathbf{o}_t)$, which specifies the optimal action (with respect to expected resulting utility) to take after observing the patient at time t. To handle the large number of observations, we use a Monte Carlo decision tree method in which predicted observations are sampled according to their likelihood. To handle the large number of actions, we preprocess the action space by eliminating options that are rarely observed in the dataset.

A reward function $R(\mathbf{o}_1, \mathbf{o}_2, \mathbf{o}_3, \mathbf{a}_1, \mathbf{a}_2)$ must be defined by the system designer over complete action/observation traces. Reward is a numerical value that may be a complex function including penalties for significant intermediate medical events, financial costs to patient or hospital, long term dependent care costs, quality of life, etc. Typically they are weighted combinations of costs of treatments and patient outcomes, as chosen by a human supervisor.

Optimizing utility. As usual in Markov Decision Processes, the optimal *utility* at different stages of the algorithm can be defined recursively from the "bottom up" as follows:

$$U(\mathbf{h}_3, \mathbf{o}_3) = R(\mathbf{o}_1, \mathbf{o}_2, \mathbf{o}_3, \mathbf{a}_1, \mathbf{a}_2) \tag{3.2}$$

$$U(\mathbf{h}_t) = \sum_{\mathbf{o}_t \in \mathcal{O}_t} U(\mathbf{h}_t, \mathbf{o}_t) P(\mathbf{o}_t \,|\, \mathbf{h}_t) \text{ for } t = 2, 3 \tag{3.3}$$

$$U(\mathbf{h}_t, \mathbf{o}_t) = \max_{\mathbf{a}_t \in \mathcal{A}_t} U(\mathbf{h}_t, \mathbf{o}_t, \mathbf{a}_t) \text{ for } t = 1, 2 \tag{3.4}$$

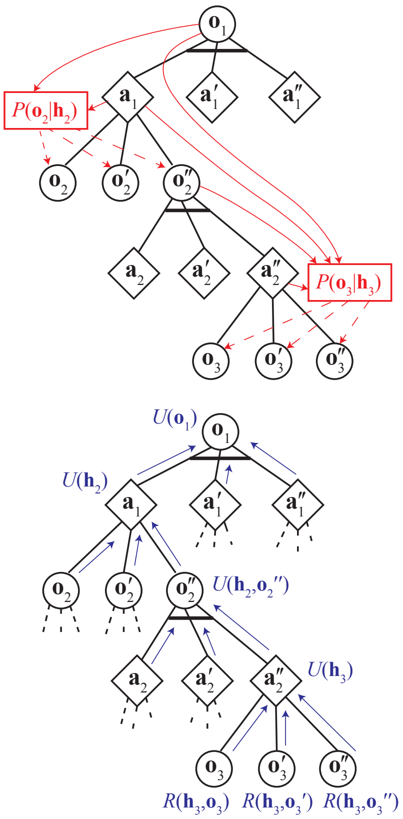


Figure 2: Computing optimal policies. Top: a decision tree is grown from the start observation by enumerating actions and sampling future observations from the treatment progression models. Bottom: rewards are backed up from leaf nodes, maximizing the utility of children at observation nodes and averaging it at action nodes. The action that yields the optimal utility at the root is chosen as the final recommendation.

where $\mathcal{O}_t$ and $\mathcal{A}_t$ are, respectively, the sets of all possible observations $\mathbf{o}_t$ and actions $\mathbf{a}_t$ at time t. The optimal action $\pi(\mathbf{h}_t, \mathbf{o}_t)$ is given by the arg max of the corresponding maximization in (3.4).

Fig. 2 illustrates the process of computing the utility function. First, it explores a game tree down to the $T = 3$ horizon, computes the utilities at the leaf nodes using the base case (3.2), and then recursively propagates backward using (3.3) and (3.4). Since it is intractable to enumerate all of $\mathcal{O}_t$, we resort to sampling $\mathcal{O}_t$ proportionally to $P(\mathbf{o}_t \,|\, \mathbf{h}_t)$ in (3.3) in a manner similar to Monte Carlo integration.

Furthermore, $\mathcal{A}_t$ is quite large, containing all combinations of individual action variables $a_t^1, \ldots, a_t^{n_t}$. However, the joint actions $\mathbf{a}_t$ observed in the dataset exhibits only a small fraction of $\mathcal{A}_t$, leading predictions based on rare or non clinically-relevant actions to have low statistical significance. For example, at the second visit, there are 6 within-study treatments and 10 potential out-of-study treatments. It would have been impractical to consider all $6 \cdot 2^{10}$ possible combinations of these. However, there are only about 150 combinations (of the 10 out-of-study treatments) that actually occur in the dataset, and only 11 that occur >0.8% of the time. Hence, we precompute the set of treatment actions $\tilde{\mathcal{A}}_t$ that are seen at least that often in the dataset, and maximize over $\tilde{\mathcal{A}}_t$ instead. The resulting $\tilde{\mathcal{A}}_2$ contains 66 actions.

Reward function. For proof-of-concept, we use a hypothetical reward function that roughly corresponds to reasonable tradeoffs between patient outcomes and costs (negative rewards).

- Each short-term administration of medicine (either one-time or on a 2-week regimen) was assumed to have a cost of 0.1, and long-term regimens (6 months) were assumed to have cost 1.
- Carotid surgery was assumed to have cost 1.
- Significant medical events were assigned costs depending roughly on their severity, either 1 (the occurrence of a minor or unspecified side effect), 2 (for more serious events like a pulmonary embolism), or 3 (for major events like recurrence of stroke).
- Death by 14 days was assigned a cost of 10.
- The final condition was rewarded as follows: −20 for death, +5 for being alive but dependent, +10 for being not completely recovered, and +20 for complete recovery.

Future implementations might use true monetary costs for treatments, expert knowledge, and established metrics like QALY to define rewards with greater relevance to real-world factors for decision-making.

4 Evaluation and Results

4.1 Data Preprocessing All trials were performed on a subset of the full IST dataset data in which:

1. All patients were confirmed to have had an ischemic stroke by CT scan prior to entry into the trial, as recommended by the IST study authors [17].
2. None of the patients were known to be noncompliant with the procedure of the study.
3. The final condition of each patient is not missing.

Out of 19,435 subjects in the original study, 12,151 were found to meet these inclusion criteria.

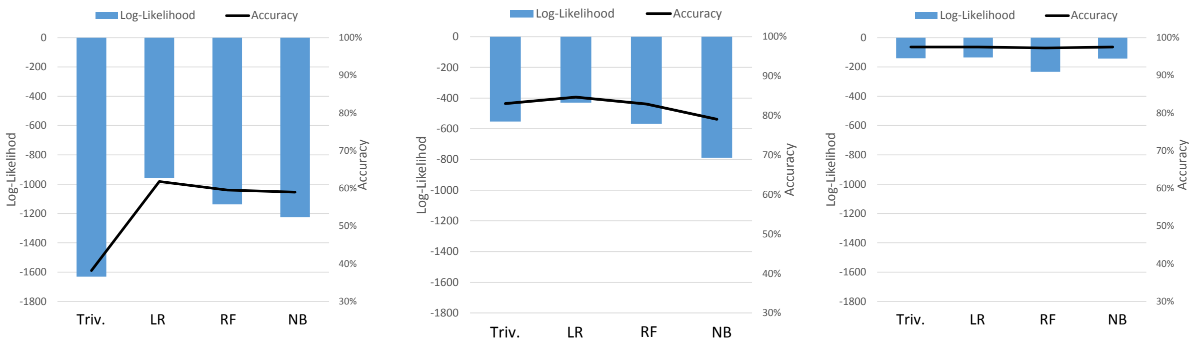
Some variables were combined in order to simplify interpretation of results. For example, the two original study variables denoting whether a patient was treated with aspirin or not and whether a patient was treated with low-dose heparin, high-dose heparin, or neither were combined into one variable with six possible values.

For LR, categorical variables were converted to indicator variables. Surprisingly we found that NB performed consistently better with the indicator rather than a categorical encoding, so the same was done for NB. For the RF, categorical variables were encoded as integer values ≥ 0, with missing values encoded as -1. This has the effect of causing the decision tree training algorithm to split based on less-than-equal/greater-than rather than equals/not-equals splits, but our experience suggests that the effect on performance is insignificant.

4.2 Model Selection Model selection was performed with respect to the cross-validated log-likelihood score on the prediction tasks $P(\mathbf{o}_3|\mathbf{h}_3)$ and $P(\mathbf{o}_2|\mathbf{h}_2)$. Log-likelihood is the appropriate metric for conditional probability estimation because it measures both the accuracy of the prediction as well as the confidence in that prediction. We also compared the accuracy of the classifier obtained by selecting the most-likely prediction. This helps understand the strengths and weaknesses of each model in performing conditional probability estimation (with log-likelihood as the appropriate metric) rather than classification (with accuracy as the appropriate metric). Overall, there were a total of 20 prediction tasks (predicting final outcome from 1st and 2nd stage variables, and then one for predicting each of the 19 intermediate targets at the 2nd stage from the 1st stage variables). For each of these tasks, models were learned and their performance upon both the log-likelihood and accuracy metrics were evaluated using 10-fold cross-validation.

First, preliminary trials yielded a surprising observation that BN consistently performed far worse than the other methods, including Trivial. The learned structure performed better than the hand-coded one, but was still outperformed by Trivial. We hypothesize that this is due to excessive data fragmentation and overfitting, causing the learned conditional probabilities to be poorly estimated. Moreover, its training time was higher than other algorithms. As a result we eliminated BN from future consideration.

Fig. 3a shows results for the task of predicting the final condition $P(\mathbf{o}_3|\mathbf{h}_3)$, which is arguably the most important model, and also the most difficult due to the large number of predictors and its 4-class multinomial domain. We see that while LR, RF, and NB significantly outperform the Trivial model, LR stands out as the clear victor, both in terms of accuracy (61.8% for LR vs. 59.6% for RF, 59.0% for NB, and 38.2% for the trivial



(a) Patient final condition

(b) Alive within 14 days

(c) Unspecified side effect within 14 days

Figure 3: Cross-validated log-likelihood and accuracy for final condition and two intermediate variables. Larger values are better.

model) and log-likelihood.

Figs. 3b and 3c show results for two representative intermediate targets: patient alive at 14 days, and occurrence of an unspecified side effect. Both are binary, and hence have higher accuracy than final condition. In Fig. 3b, LR is still clearly the best performer on both metrics, with an accuracy of 84.68% (as compared to 82.9% for NB, 79.1% for RF, and 83.1% for the trivial model). The NB model does reasonably well in terms of accuracy, but has a lower log-likelihood; this is because its estimates of posterior class probabilities are heavily biased due to its strong independence assumptions. In Fig. 3c, the occurrence of a side effect is rare, and none of the models manage to significantly outperform the trivial model. Notice that NB and RF have switched rank relative to each other in Fig. 3c, but LR still outperforms both.

We found that LR consistently outperformed or did no worse than the other learners on all other intermediate targets. As a result, we used LR to learn the treatment progression models for the CDSS.

4.3 MDP Performance This section evaluates the performance of the CDSS optimized policies on a 1,215 patient holdout set. Except when otherwise noted, these experiments used 5 observation samples per layer of the MDP decision tree.

Comparison against baselines. First, we compared the utility of the MDP optimized policies (MDP) against two baselines: 1) Actual: the decisions made by actual human doctors in the dataset (including the random assignment to the aspirin / heparin treatment groups), and 2) Greedy: greedy 1-step lookahead. The greedy strategy is equivalent to a rollout of the MDP decision tree limited to depth 2 (i.e, one action and one predicted observation), and summation of utilities in (3.3) are approximated using 20 Monte Carlo observation samples. Fig. 4 plots the final mean utility values across the holdout set, showing that greedy optimization performs poorly, while the MDP optimization produces higher utilities than the actual decisions made by the human doctors. The standard deviation of utility in all models was approximately 16. Statistical significance testing indicates the difference between Actual and Greedy is insignificant ($p > 0.05$), and the difference between Actual and MDP is significant ($p < 0.001$).

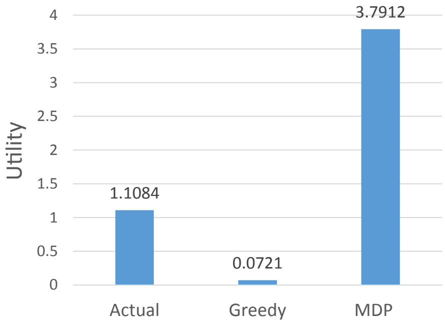


Figure 4: Utilities for the actual policy used by human doctors, a greedy policy, and the MDP policy. Larger values are better.

Fig. 5 compares the simulated distributions of the 6-month followup condition for each policy. Results suggests that the MDP policy may reduce a patient's risk of death by 4% and increase the chance of fully recovering by 11% (in absolute terms). These differences are significant ($p < 0.001$). The MDP policy also reduced the simulated risk of dying within the first 14-day followup by 2% but this reduction was insignificant ($p > 0.05$).

Computation time. We studied how the number of observation samples affects CDSS utility and computa-

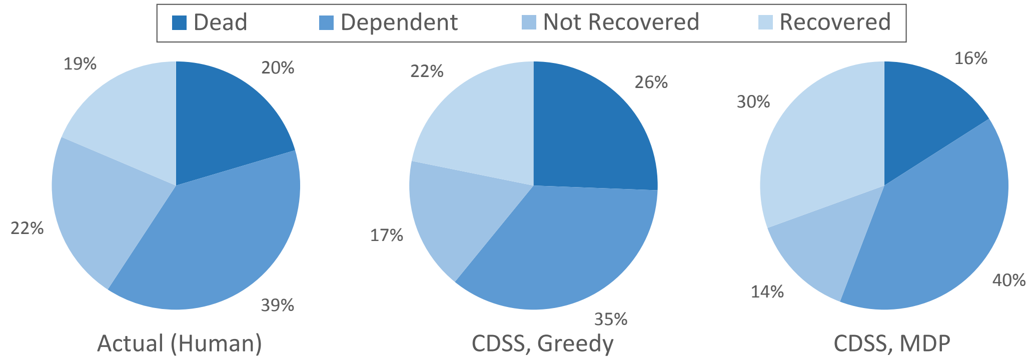


Figure 5: Distribution of the final patient condition after 6 month followup, actual (left) and a greedy policy, and the MDP CDSS policy.

tion time. We note that computation speed is not a major issue in this domain due to the small lookahead; with 5 observation samples, the CDSS produces optimized policies in approximately 13 s on a 3.4 GHz Intel Core i7 CPU. For 1 sample, computation took 1.3 s; for 2 samples, 5.0 s; for 8 samples, 37 s; and for 10 samples, 55 s. The CDSS is implemented in unoptimized Python code, and for longer-horizon or larger MDPs our approach would benefit from implementation in a compiled language.

Policy quality suffers only at very small sample sizes. Increasing sample count from 1 to 5, the resulting simulated utilities were 0.67, 1.9, 2.4, 3.7, and 3.6. Surprisingly it appears that 5 samples is sufficient to get reasonably good policies; at 4 samples and above, policy quality did not change significantly.

Comparison with clinical practices. Lastly, we examined how CDSS decisions differ from those of human analysts and clinical practice. First, we remark that a prior analysis of IST indicates inconclusive results regarding the use of aspirin and heparin, with a small but weakly significant or insignificant improvements in survival rate [14]. Double-blind clinical studies on a smaller cohort indicate little additional benefit to administration of heparin compared to aspirin [5].

Using the CDSS, the suggested rates of pretrial use of aspirin and/or heparin — that is, administering medication as soon as possible after stroke diagnosis — jumped from 22% to 58% of patients. This is consistent with current best practices that recommend administration of medication immediately after stroke [6].

Compared to the initial randomized administration of medicine, the CDSS overwhelmingly preferred combining aspirin and low-dose heparin, which was administered to 66% of test patients. Although current practice advises against the use of heparin due to the increased risk of hemorrhage [14], our CDSS picks up patterns in the dataset suggesting the combination of aspirin and low-dose heparin does have a positive effect beyond aspirin alone. Comparing the cohort in our test set of patients receiving aspirin + low-dose heparin vs aspirin alone in IST, we observe a slightly *reduced* rate of cerebral hemorrhage (0.7% vs 0.9%), an insignificant change in rates of stroke reoccurrence (3.0% vs 3.1%), increased risk of unspecified side effect (3.3% vs 1.3%), and an increased survival rate (79% vs 76%). These differences are not statistically significant. In any case, the CDSS prefers aspirin + low-dose heparin because it is given larger rewards for survival over the costs of side-effects.

The CDSS also administered more treatments outside the scope of the IST randomization, with significant differences in the rates of use of subcutaneous heparin (7.6% vs 2.7%), intravenous heparin (8.4% vs 2.9%), other anticoagulants (14.9% vs 2.8%), calcium antagonists (26.6% vs 12.4%), and haemodilution (21.4% vs 4.4%). This increased treatment rate may be due to a direct causal relation, but we find that the literature does not support such an argument [2]. Instead, it could be an effect of a hidden influence, such as hospital quality. The reasoning is that better-equipped hospitals may have more capacity to provide additional treatments and may also have higher survival rates in general.

5 Conclusion

We have described a general purpose CDSS for multistep conditions that learns probabilistic temporal models from complex clinical trial data and applies them in an MDP framework. Applied to the International Stroke Trial dataset, we applied model selection to select the learning technique that optimizes predictive accuracy. The resulting optimized policies increase the survival rate and full recovery rate significantly.

This work has some technical limitations. One is that we make a simplifying assumption that intermediate observables are conditionally independent, given history. It may be possible to relax this assumption, for example, by learning conditional random field models. The MDP solved by our CDSS is also moderately-sized, and better solution techniques may be needed to scale to much larger domains with dozens of time points and thousands of observables and actions.

Big-picture limitations also remain, which should prove to be rich areas for future research. First, clinical data will never observe multiple courses of treatment applied to the same person, so it is difficult to confidently make a recommendation for rare but promising treatments — a problem faced by human doctors as well. To address this, a CDSS might indicate its level of certainty to the clinician using confidence intervals, and recommend decisions that are robust to modeling errors. Second, recommendations are only as good as the data (i.e., garbage-in, garbage-out). Without a "ground truth" model for evaluation, we cannot confidently claim that our system is truly making clinically superior decisions. In the near future, we plan to present the system's chain of reasoning to human domain experts to ensure that the recommendations are, at the very least, logical. If they pass this test, then we hope to ultimately evaluate performance of our CDSS in randomized clinical trials.

References

[1] O. Alagoz, H. Hsu, A. Schaefer, and M. Roberts. Markov decision processes: A tool for sequential decision making under uncertainty. *Medical Decision Making*, 30(4):474–483, 2009.

[2] K. Asplund. Haemodilution for acute ischaemic stroke. *Cochrane Database Syst Rev*, (4):CD000103, 2002.

[3] D. Bates, G. Kuperman, S. Wang, A. K. T. Gandhi, L. Volk, C. Spurr, R. Khorasani, M. Tanasijevic, and B. Middleton. Ten commandments for effective clinical decision support: Making the practice of evidence-based medicine a reality. *Journal of the American Medical Informatics Association*, 10(6):523–530, 2003.

[4] C. Bennett and K. Hauser. Artificial intelligence framework for simulating clinical decision-making: a markov decision process approach. *Artificial Intelligence in Medicine*, 57(1):9–19, 2013.

[5] E. Berge, M. Abdelnoor, P. Nakstad, and P. Sandset. Low molecular-weight heparin versus aspirin in patients with acute ischaemic stroke and atrial fibrillation: a double-blind randomised study. *The Lancet*, 355(9211):1205–1210, 2000.

[6] Z. Chen, P. Sandercock, H. Pan, C. Counsell, R. Collins, L. Liu, J. Xie, C. Warlow, and R. Peto. Indications for early aspirin use in acute ischemic stroke: A combined analysis of 40,000 randomized paients from the chinese acute stroke trial and the international stroke trial. *Stroke*, 31(6):1240–1249, 2000.

[7] K. Davis, K. Stremikis, D. Squires, and C. Shoen. Mirror, mirror on the wall: How the performance of the U.S. health care system compares internationally (2014 update). Technical Report 1755, The Commonwealth Fund, 2014.

[8] J. Eom, S. Kim, and B. Zhang. Aptacdss-e: A classifier ensemble-based clinical decision support system for cardiovascular disease level prediction. *Expert Systems with Applications*, 34(4):2465–2479, 2008.

[9] J. Goulionis and A. Vozikis. Medical decision making for patients with parkinson disease under average cost criterion. *Australia and New Zealand Health Policy*, 6(15), 2009.

[10] R. Kaushal, K. Shojania, and D. Bates. Effects of computerized physician order entry and clinical decision support systems on medication safety: A systematic review. *Archives of Internal Medicine*, 163(12):1409–1416, 2003.

[11] K. Kawamoto, C. Houlihan, E. Balas, and D. Lobach. Improving clinical practice using clinical decision support systems: A systematic review of trials to identify features critical to success. *BMJ*, 330(7494):765, 2005.

[12] M. Kearns, Y. Mansour, and A. Ng. A sparse sampling algorithm for near-optimal planning in large markov decision processes. *Machine Learning*, 49(2-3):193–208, 2002.

[13] M. Littman and R. Sutton. Predictive representations of state. In *Advances in Neural Information Processing Systems*, pages 1555–1561, 2002.

[14] S. Ricci, S. Lewis, P. Sandercock, and IST Collaborative Group. Previous use of aspirin and baseline stroke severity: An analysis of 17,850 patients in the international stroke trial. *Stroke*, 37(7):1737–1740, 2006.

[15] R. Sacco, P. Wolf, W. Kannel, and P. McNamara. Survival and recurrence following stroke. the framingham study. *Stroke*, 13(3):290–5, 1982.

[16] P. Sandercock, R. Collins, C. Counsell, B. Farrell, R. Peto, J. Slattery, C. Warlow, and IST Collaborative Group. The international stroke trial (IST): A randomised trial of aspirin, subcutaneous heparin, both, or neither among 19 435 patients with acute ischaemic stroke. *The Lancet*, 349(9065), 1997.

[17] P. Sandercock, M. Niewada, A. Czonkowska, and IST Collaborative Group. The international stroke trial database. *Trials*, 12(101), 2011.

[18] A. Towfighi and J. Saver. Stroke declines from third to fourth leading cause of death in the united states: Historical perspective and challenges ahead. *Stroke*, 42(8):2351–2355, 2011.

[19] M. Verduijn, P. Rosseel, N. Peek, E. de Jonge, and B. de Mol. Prognostic bayesian networks: I: Rationale, learning procedure, and clinical use. *Journal of Biomedical Informatics*, 40(6):609 – 618, 2007.

[20] M. Verduijn, P. Rosseel, N. Peek, E. de Jonge, and B. de Mol. Prognostic bayesian networks: II: An application in the domain of cardiac surgery. *Journal of Biomedical Informatics*, 40(6):619 – 630, 2007.