

# Spectral Machine Learning for Predicting Power Wheelchair Exercise Compliance

Robert Fisher<sup>1</sup>, Reid Simmons<sup>1</sup>, Cheng-Shiu Chung<sup>2</sup>, Rory Cooper<sup>2</sup>, Garrett Grindle<sup>2</sup>, Annmarie Kelleher<sup>2</sup>, Hsinyi Liu<sup>2</sup>, and Yu Kuang Wu<sup>2</sup>

<sup>1</sup>Carnegie Mellon University  
5000 Forbes Ave, Pittsburgh PA 15213

<sup>2</sup>Human Engineering Research Laboratories  
University of Pittsburgh  
6425 Penn Avenue, Pittsburgh PA 15206

**Abstract.** Pressure ulcers are a common and devastating condition faced by users of power wheelchairs. However, proper use of power wheelchair tilt and recline functions can alleviate pressure and reduce the risk of ulcer occurrence. In this work, we show that when using data from a sensor instrumented power wheelchair, we are able to predict with an average accuracy of 92% whether a subject will successfully complete a repositioning exercise when prompted. We present two models of compliance prediction. The first, a spectral Hidden Markov Model, uses fast, optimal optimization techniques to train a sequential classifier. The second, a decision tree using information gain, is computationally efficient and produces an output that is easy for clinicians and wheelchair users to understand. These prediction algorithms will be a key component in an intelligent reminding system that will prompt users to complete a repositioning exercise only in contexts in which the user is most likely to comply.

**Keywords:** Machine learning, spectral learning, HMMs, healthcare applications

## 1 Introduction

More than 2.5 million Americans experience pressure ulcers every year[14]. One particularly high risk group of individuals are those with severe physical disabilities that utilize power wheelchairs as their primary means of mobility. Power wheelchairs can very effectively relieve pressure by raising the user to an reclined, titled position for several seconds. However, research indicates that less than 40% of power wheelchair users correctly use their power seat functions to relieve pressure and prevent deadly ulcers[12]. This leads many power wheelchair users to be exposed to preventable pressure ulcer formation, which can often lead to complicating infections or even death.

The Virtual Coach smart power wheelchair system was designed to track the power wheelchair usage of users in order to help them better conform to

the pressure relief guidelines set forth for them by clinicians, as well as improve their overall posture. By monitoring encoders in the power wheelchair’s joints (chair tilt, leg rest elevation, seat elevation, etc), the system is able to determine when a user has successfully performed the repositioning exercise prescribed by the user’s healthcare provider. Users are reminded on a periodic basis to conduct pressure relief, and the system tracks whether the user complied with the reminder or not. Initial results suggest that users that receive regular reminders from the system have higher rates of compliance with their exercise regime than those users not receiving instruction. However, there is a risk of the system becoming intrusive and annoying if users are given too many undesired reminders.

Instead of simply reminding users with a static periodicity, we would like to devise an intelligent system that selects the most appropriate contexts in which to issue reminders. In this paper, we present the results of applying machine learning algorithms to predict if a user is likely to comply with a reminder given information about the user’s current context collected by several sensors onboard the Virtual Coach system. We show that when averaged across all users, we can attain a predictive accuracy of 92%. We present results with an expressive and statistically consistent spectral Hidden Markov Model, as well as a computationally lightweight decision tree model that creates an output which is easy for clinicians and users to understand. The goal for this work is to create an intelligent, decision theoretic reminding system that would select the best moments to issue reminders to users given the likelihood of compliance in the current context, the time since the last pressure relief was completed, and the recommendations given by the overseeing clinician. Predicting whether a user is likely to comply in a given context is the first step towards realizing these goals.

## 2 Related Work

Predicting pressure relief compliance using sensor data is related to the task of predicting user interruptability given contextual information. For instance, there has been some work predicting whether or not a user would prefer to have their smartphone ringer enabled in a given setting. Some of this work has leveraged decision theoretic models of user preferences[15], while other work has emphasized active sampling of user preferences coupled with features generated using sensor processing algorithms[10].

There has been some previous work utilizing machine learning to predict patient behavior in healthcare applications, but rarely with the fine level of granularity seen in the Virtual Coach system. For instance, one author used statistical machine learning techniques with patient data to predict which subjects suffering from coronary artery disease would be likely to comply with pharmaceutical guidelines for managing cholesterol levels[8]. These predictions were based primarily on demographic data, and did not give indications as to which contexts and circumstances would lead to non-compliance for a subject. There has also been quite a bit of work using machine learning to predict health care outcomes

and complications, but this work has not been focused on understanding or altering patient behavior to improve outcomes[16, 6].

In this paper, we present results using *spectral methods* for machine learning, which has become a very popular topic in machine learning research. Based on eigenvector decomposition, such as is used in Singular Value Decomposition (SVD) and Principle Component Analysis (PCA), spectral methods can give optimal results for many optimization tasks using a small fraction of the computation required by comparable algorithms. Spectral algorithms exist for learning latent-variable PCFG's[5], dynamical systems[3], and Hidden Markov Models[11]. From an applications standpoint, spectral methods have been used to estimate a student's aptitude for key tasks in a classroom setting[9], for dependency parsing of natural language text[7], and for image segmentation and classification tasks[13].

### 3 The Virtual Seating Coach System

The Virtual Seating Coach (VC) system is a smart power wheelchair outfitted with a variety of sensors and an onboard computer. The sensors installed on these chairs include encoders in each of the chair's joints and wheels, accelerometers in the base of the seat, a seat occupancy sensor, a thermometer, and a light sensor. Additionally, a subset of users in the clinical trails used chairs that were outfitted with GPS chips and microphones. A complete list of the sensors and the machine learning features that were computed is shown in Table 1.



**Fig. 1.** A researcher demonstrating the use of the Virtual Coach system

A tablet computer is attached to the arm of the chair, as seen in Figure 1. The software on this tablet is used to issue reminders to users, and to provide feedback while an exercise is being conducted. Additionally, users can set preferences that dictates if reminders will be given using audio queues, silent text, or if the reminders should be disabled completely. When a reminder is given, a user is

given the choice of snoozing the reminder, which will result in another prompt being given in 5 minutes, or the reminder can be dismissed for 1 hour. If the user does not comply with the reminder, it is automatically snoozed for 5 minutes. If the user successfully completes the exercise, the reminding system will be reset for a duration determined by a clinician.

Data Source	Features
Chair angle encoders	Mean and variance of encoder values over previous 10 seconds
Wheel Encoders	Mean and variance of rotational velocity over previous 10 seconds
Accelerometers	Average Fourier power in 500 Hertz bands over 0-4000 Hz
Thermometer	Average temperature over previous 30 seconds
Seat Occupancy	A 0/1 value indicating state of pressure sensor in seat
Clock	The current day of the week and hour of the day
Recent Behavior	Time since the last reminder
Recent Behavior	Outcome of previous reminder (Snooze, dismiss, compliance)
GPS Data <sup>1</sup>	The current encrypted longitude and latitude coordinates
Audio Data <sup>1</sup>	Average Fourier power in 100 Hertz bands over 0-1000 Hz

Table 1: The features computed using the Virtual Coach sensors

Several prototype versions of the system have been undergoing clinical trials, and more than 20 volunteers have participated in the trials using the VC system over a period of 6 weeks per user. Volunteers were either assigned to a control group, being given a sensor equipped chair without a reminding system, or they were given the full VC system with tablet computer. The system stores clinician pressure relief guidelines for the individual user. For instance, the recommendation may be for the user to receive 30 seconds of pressure relief once every 60 minutes. During clinical trials the system issued reminders on a fixed basis according to the prescription, and the user’s response to the reminder was recorded.

## 4 Spectral Hidden Markov Models

In recent years, spectral methods have become increasingly popular tools for solving a wide variety of optimization problems in machine learning. Spectral algorithms utilize eigenvector decomposition of data matrices to estimate the parameters of a learning model. Spectral algorithms are statistically consistent, do not suffer from local optima, and can be orders of magnitude faster than comparable optimization techniques, such as Expectation Maximization (EM). Spectral algorithms have been particularly successful when learning latent variable models[11].

Hidden Markov Models operate on sequences of observed data and assume the existence of a latent state variable that allows for the Markovian assumption. Put another way, given a hidden state  $h_i \in [1, m]$  and a sequence of observations

<sup>1</sup> Data only available for subset of users

$x_1, x_2, \dots, x_t, x_i \in \mathbb{R}^n$ , the assumptions driving the HMM give us the following equality:

$$P(x_t|x_1, x_2, \dots, x_{t-1}) = P(x_t|h_i)$$

Because the hidden state is unobserved and latent, the current state is generally represented as a distribution over possible states. We can then marginalize over the latent state distribution to estimate the probability of observing a given outcome.

There are three sets of parameters that must be learned to describe an HMM. The state transition probability matrix,  $T \in \mathbb{R}^{m \times m}$ , defines the probability of transitioning from one hidden state to another. The observation matrix,  $O \in \mathbb{R}^{n \times m}$ , defines the probability of seeing a given observation given a state, where  $n$  is the dimensionality of our observation space. Finally, the initial state distribution,  $\pi \in \mathbb{R}^m$ , defines the probability of beginning in each of the  $m$  latent states. The mean squared error of estimating  $T$  and  $O$  goes to 0 with the number of observations. However, the mean squared error of the estimate of  $\pi$  goes to 0 with the number of observed sequences. Therefore, a model trained on a single sequence of many observations may have high predictive error because the estimate of  $\pi$  will be incorrect. We prefer, therefore, to use several short independent sequences of operations. For our application, we consider each day that a subject uses the Virtual Coach system to be a separate independent sequence. In practice it is not necessary to know  $m$ , the number of hidden states, because this value can be tuned using cross validation.

Historically, the parameters of an HMM have been trained using the EM algorithm or a similar approach. However, these optimization techniques tend to be very slow and prone to fall into local optima. This is in part due to the fact that learning HMM parameters exactly has been shown to be intractable under cryptographic assumptions[17]. Recently, spectral formulations of the HMM learning problem have been posed which allow an HMM to be learned quickly and optimally when the parameter matrices are of rank no greater than  $m$ [1, 11].

The spectral formulation of an HMM utilizes subspace identification by creating an  $m$ -dimensional subspace that preserves the state transition dynamics of the original model. This type of subspace identification is closely related to the use of Predictive State Representations, which have been used quite successfully in a variety of dynamical systems applications[4]. To project into this subspace, we must compute an invertible subspace transformation matrix,  $U \in \mathbb{R}^{n \times m}$ , that allows us to transform data in the original  $n \times n$  space into the  $n \times m$  subspace. To accomplish this, we begin by computing the empirical probabilities of sequential cooccurrence of all observation doubles and store the results in a matrix  $\hat{P}_{2,1} \in \mathbb{R}^{n \times n}$ . Put another way, this matrix represents the probability of seeing observation  $x_i$  followed immediately by observation  $x_j$ . The matrix  $U$  can then be recovered by performing Singular Value Decomposition (SVD) on the matrix  $\hat{P}_{2,1}$ . Now, rather than learning the HMM parameters in the original space, we learn the parameters in this reduced subspace. If the subspace dimensionality,  $m$ , is no less than the rank of the parameter matrices  $T$  and  $O$ ,

then this subspace embedding will act as a lossless compression, and the new parameters that we learn will have the same dynamics as the original system. Of further interest, these new subspace parameters can be computed using only computationally efficient operations, such as empirical probability estimates and standard matrix operations. We denote the subspace parameters as  $\hat{\pi}_U$ ,  $\hat{T}_U$ , and  $\hat{O}_U$ .  $X^+$  represents the Moore-Penrose pseudo-inverse of a matrix.  $P_1$ ,  $P_{2,1}$ , and  $P_{3,x,1}$  respectively represent the empirical single, double, and triple cooccurrence probabilities of observation. It can then be proven that the following equalities hold:

$$\begin{aligned}\hat{\pi}_U &= U^T P_1 \\ \hat{T}_U &= (P_{2,1}^T U) + P_1 \\ \hat{O}_U &= U^T P_{3,x,1} (U^T P_{2,1})^+ \forall x \in [n]\end{aligned}$$

For the sake of brevity, we omit the proofs here, but they can be found in Hsu, *et al* 2012[11]. We see from the above equalities that the parameters of the spectral HMM can be recovered exactly using only empirical distribution estimates with matrix transposition, multiplication, and pseudo-inverse operations.

## 5 Decision Trees

Power wheelchairs are a principle component in the life of many people living with a physical disability, and research suggests that chair users prefer to remain in full control of their chair’s functionality [2]. In light of this, we may wish to create a reminding system that is maximally transparent, so that a user can better understand how the intelligent system is deciding when to remind. The spectral HMM, while very powerful, relies on a latent variable distribution that often lacks a simple, relatable explanation.

In an effort to explore a model that would be easier for end users to comprehend, we turned to one of the oldest and most popular machine learning classifiers: decision trees. Decision trees conduct classification through a series of logical binary operations applied to the learning features. *Information gain* is most often used to determine which features to place near the root of the tree. Information gain is defined as the reduction in statistical entropy gained by learning the state of a random variable. If we denote  $H(Y)$  to be the statistical entropy of a random variable, the information gain of  $X$  with regards to label  $Y$  is  $H(Y) - H(Y|X = a)$ .

Decision trees have been shown to be high bias classifiers, leading to a great deal of overfitting [18]. To combat this, decision trees are generally pruned near the leaves of the tree to reduce bias and improve the tree’s capacity for generalization. Smaller trees will also be easier for users to view and comprehend.

An example of a heavily pruned decision tree trained on one user’s Virtual Coach data is shown in Figure 2. In this figure, a lead node of 1 indicates that the user is expected to comply with a reminder, 0 indicates predicted non-compliance. The tree in this figure was trimmed to depth 3, leaving only encrypted location and audio features remaining. It is worth noting that the audio

feature with the highest information gain is the Fourier signal power in the 300 to 400Hz range of the spectrum. This represents the lower end of the audio spectrum that the human voice inhabits, so the decision tree seems to indicate that this given user would prefer not to be reminded if there is evidence of a speech signal present in the environment. If a Virtual Coach user were to elect to use a decision tree to determine when the intelligent system should issue reminders, the user would be able to verify, and possibly modify, the behavior of the system. This would help to curb any confusion a user may have as to why the system is behaving in the way that it is.

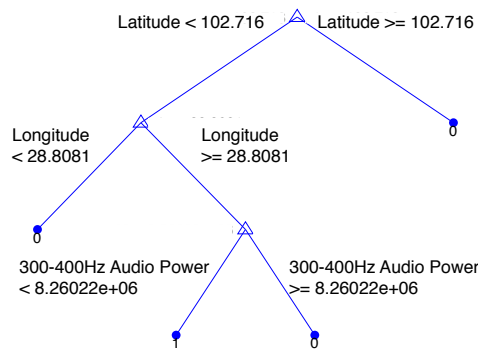
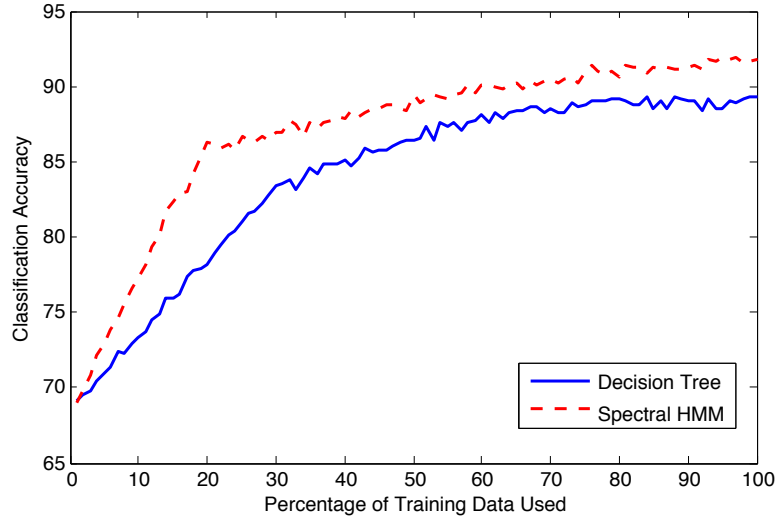


Fig. 2. A Decision Tree Showing the Compliance Predictions of a User

## 6 Results

When collecting empirical results, we used 15 fold cross validation to determine prediction accuracy of our models by randomly selecting one week of data to hold out for validation. Figure 3 shows the average prediction accuracy of the spectral HMM and decision tree models as a function of training data volume. We also considered several other classifiers, including Support Vector Machines, Nearest Neighbors, Naive Bayes' algorithm, and Conditional Random Fields. None of these other approaches performed as well as the spectral HMM, or produced output that was as easy to understand as the decision trees.

We see from these results that the spectral HMM performs particularly well with small amounts of data, with a very steep peak in the training curve after being supplied with only 20% of the data (roughly one week's worth of data). This result fits with intuition, because we expect accuracy to plateau after seeing one full instance of a user's weekly schedule. The spectral HMM peaks at around 92% classification accuracy, while the decision tree model peaks at 89%. It is not



**Fig. 3.** Empirical Training Curve

surprising that the spectral HMM performs the best, because the HMM reflects recent observations through the latent state distribution, while the decision tree performs classification using only the current sensor values. For comparison, a naive model that always predicts the most likely prior attains 69% classification accuracy.

We determined that both models performed best when supplied with training and testing data from only a single user. Models trained with data from multiple users resulted in many contradictory training examples, which is intuitive, because we would expect different users to react differently to a reminder under the same circumstances.

Prediction accuracy with users for whom audio and GPS data were available was higher than the rest of the population, showing 94% average accuracy with the HMM, and 92% average accuracy with the decision tree. We see a smaller differential between the two models in this case due to the fact that the GPS data shows very high information gain, often placing it near the root of the decision trees. For these users we can attain a predictive accuracy surpassing 85% using trees of only depth four that were built only using time of day, GPS coordinates, and the outcome of the most recent reminder. This suggests that simpler models are almost as effective for users that are willing to share personal information, such as location data, with the Virtual Coaching system.



## 7 Conclusions and Future Work

Initial results show that we can accurately predict repositioning exercise compliance given a set of fairly common sensors embedded in a power wheelchair. The spectral Hidden Markov Model leverages Markovian information about the past in a framework that has many desirable theoretical and practical qualities. On the other hand, the decision tree model is computationally very lightweight, and presents an easily understandable graphical representation of the model. The tradeoffs of these two methods means that each may be preferred to the other under certain conditions.

We are currently in the process of designing a new Virtual Coach system that runs on a user's smartphone and requires very little effort to install the system in a commercial power wheelchair. We are working to incorporate the predictive capabilities presented in this paper in an intelligent reminding system to be made available to end users. Some additional work is necessary to use a prediction algorithm in an intelligent reminding system. First, the intelligent system must balance an exploration/exploitation trade-off when deciding when to remind. For instance, a user may be in a context in which they are likely, but not guaranteed, to respond to a reminder, but we may find more optimal conditions for reminding in 10 minutes time. Questions related to the human interface design of the intelligent system must also be considered to guarantee that users continue to feel in control of the technology when it is making decisions.

By using machine learning to monitor and alter users' pressure relief habits, we have the potential to prevent thousands of occurrences of pressure ulcers every year. The cost saving nature of this technology will also make it desirable for insurance companies and healthcare organizations, which could lead to faster adoption of the technology and improved outcomes for the millions of power wheelchair users in the world.

## 8 Acknowledgements

This material is based upon work supported by the National Science Foundation under Cooperative Agreement EEC-0540865 as well as by a National Science Foundation Graduate Research Fellowship. We also acknowledge the Pittsburgh chapter of the American Rewards for College Scientists (ARCS) program for their generous support.

## References

1. Raphael Bailly. Quadratic weighted automata: Spectral algorithm and likelihood maximization. *Journal of Machine Learning Research*, 20:147–162, 2011.
2. Scott R Beach, Richard Schulz, Judith T Matthews, Karen Courtney, and Annette DeVito Dabbs. Preferences for technology versus human assistance and control over technology in the performance of kitchen and personal care tasks in baby boomers and older adults. *Disability and Rehabilitation: Assistive Technology*, (0):1–13, 2013.

3. Byron Boots and Geoffrey J Gordon. An online spectral learning algorithm for partially observable nonlinear dynamical systems. In *AAAI*, 2011.
4. Byron Boots, Sajid M Siddiqi, and Geoffrey J Gordon. Closing the learning-planning loop with predictive state representations. *The International Journal of Robotics Research*, 30(7):954–966, 2011.
5. Shay B Cohen, Karl Stratos, Michael Collins, Dean P Foster, and Lyle Ungar. Spectral learning of latent-variable pcfgs. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*, pages 223–231. Association for Computational Linguistics, 2012.
6. Joseph A Cruz and David S Wishart. Applications of machine learning in cancer prediction and prognosis. *Cancer informatics*, 2:59, 2006.
7. Paramveer S Dhillon, Jordan Rodu, Michael Collins, Dean P Foster, and Lyle H Ungar. Spectral dependency parsing with latent variables. In *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning*, pages 205–213. Association for Computational Linguistics, 2012.
8. Anil K Dubey. Using rough sets, neural networks, and logistic regression to predict compliance with cholesterol guidelines goals in patients with coronary artery disease. In *AMIA Annual Symposium Proceedings*, volume 2003, page 834. American Medical Informatics Association, 2003.
9. Mohammad H Falakmasir, Zachary A Pardos, Geoffrey J Gordon, and Peter Brusilovsky. A spectral learning approach to knowledge tracing. 2010.
10. Robert Fisher and Reid Simmons. Smartphone interruptibility using density-weighted uncertainty sampling with reinforcement learning. In *Machine Learning and Applications and Workshops (ICMLA), 2011 10th International Conference on*, volume 1, pages 436–441. IEEE, 2011.
11. Daniel Hsu, Sham M Kakade, and Tong Zhang. A spectral algorithm for learning hidden markov models. *Journal of Computer and System Sciences*, 78(5):1460–1480, 2012.
12. Michèle Lacoste, Rhoda Weiss-Lambrou, Magali Allard, and Jean Dansereau. Powered tilt/recline systems: why and how are they used? *Assistive Technology*, 15(1):58–68, 2003.
13. Ha Quang Minh, Marco Cristani, Alessandro Perina, and Vittorio Murino. A regularized spectral algorithm for hidden markov models with applications in computer vision. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2384–2391. IEEE, 2012.
14. Madhuri Reddy, Sudeep S Gill, and Paula A Rochon. Preventing pressure ulcers: a systematic review. *Jama*, 296(8):974–984, 2006.
15. Stephanie Rosenthal, Anind K Dey, and Manuela Veloso. Using decision-theoretic experience sampling to build personalized mobile phone interruption models. In *Pervasive Computing*, pages 170–187. Springer, 2011.
16. Xiaowei Song, Arnold Mitnitski, Jafna Cox, and Kenneth Rockwood. Comparison of machine learning techniques with classical statistical models in predicting health outcomes. *Medinfo*, 11(Pt 1):736–40, 2004.
17. Sebastiaan A Terwijn. On the learnability of hidden markov models. In *Grammatical Inference: Algorithms and Applications*, pages 261–268. Springer, 2002.
18. Allan P White and Wei Zhong Liu. Technical note: Bias in information-based measures in decision tree induction. *Machine Learning*, 15(3):321–329, 1994.