

# Health State Modeling and Prediction based on Hidden Markov Models

Kokoy Siti Komariah

Dept. of IT Convergence and Applications Engineering  
Pukyong National University  
Busan, Korea  
kokoyksk@pukyong.ac.kr

Bong-Kee Sin

Dept. of IT Convergence and Applications Engineering  
Pukyong National University  
Busan, Korea  
bkshin@pknu.ac.kr

**Abstract**—The collective health clinic data of people in a society is surmised to have a variety of characteristic health states and certain dynamics governing health state changes over time. Given such a collection of samples, we propose a way of estimating a set of health states in dynamic context using the tool of hidden Markov model (HMM). We also present a method of predicting the future health states based on the Markov dynamics or a set of state duration statistics derived from the model parameters and data clusters. In the proposed method we design a number of HMMs, each for a set of sequences with a particular disease history of interest. They are used to predict the future health states on the basis of corresponding diseases or health problems. Health state prediction as a service can be presented with a set of potential future states and over a number of years into the future. Experimental results have shown a baseline performance of 48% for single year prediction with a model of sixteen states with single hypothesis for the apoplexy history data set. The performance quickly increases to 70% with two hypotheses. When applied to multiple year prediction, the accuracy decreases by about 12 percentage points or less over the course of five years.

**Keywords**—Health clinic data, health state, hidden Markov models, data clustering, state prediction

## I. INTRODUCTION

Health clinic data from a population has the potential to provide us with a valuable insight into the collective health information which includes a variety of health states and their dynamic changes over time [1]. Health clinic data is a time series in which, unlike static images, the temporal dimension of the data carries a key aspect of the current health condition. One popular tool for modeling such a time series dynamics is hidden Markov model or HMM [2]. It has also drawn attention as convenient tool for many prediction tasks [3, 4].

There is a recent surge of interest in health clinic data analysis with practical application to health state evaluation and forecasting future development. Kawamoto et al. conducted a research on the analysis of health checkup data using the HMM [5]. Although famed as a powerful modeling tool for noisy and highly variable signals, the HMM generally falls short of the expectation. This problem is particularly serious when it comes to health state prediction tasks. Hence other types of models have since been tried for improved prediction performance [6, 7]. Still, however, the HMM has an undeniable charm as an elegant modeling tool for time series data analysis and this gives us all the more reason to explore the predictive power of the model.

Given a collection of samples, it is tempting to predict the future states by consulting the samples of people with similar health states. For this we may only have to prepare a set of clusters of relevant samples [8]. One problem of this approach is the requirement that we compare and collect similar health state trajectories; merely finding a similar state is not enough [9, 10]. Computing similarity between partial sequences of different lengths is enabled by using HMMs with which we can identify the current health state of a person in dynamic context that provides the condition from which the health state will evolve out [11, 12].

In this research we found it desirable to define a set of health states in dynamic context as manifested by a collection of samples. A person's health trajectory can start and end at any state and therefore a brute force clustering of time series data will only lead to exponentially many set of clusters. In the standard HMM approach, the state space is a finite set of  $N$  states. With an appropriate learning, each state will come to characterize a cluster of health state vectors. One advantage of using HMM is that the health state dynamics is learned simultaneously from the data [13]. Hence the proposed approach brings us a streamlined solution to health state prediction for any individual given a current state.

The remainder of the paper is organized into 5 sections. Section 2 describes the health clinic data as a time series. Section 3 presents the concept of health states and then gives a formal definition of HMM for health state and dynamics. Section 4 discusses the method of health state prediction, one-step single year and long-term prediction. Section 5 gives an extensive set of experiments to confirm the proposed method. Finally Section 6 concludes the paper.

## II. HEALTH CLINIC SEQUENCE DATA

### A. Health Clinic Data

Health clinic data comprise a sequence of records that include a number of clinical examinations and health interview answers and are prepared in a fixed format at regular intervals. Some variables are continuous valued while many others are simple choices and discrete valued. We assume that the record is given as a fixed dimensional vector of attributes. In this study the original data set consists of sequences of such records each with fifty attributes including body weight and height, blood pressure and sugar levels, cholesterol levels, and smoking, drinking and other habits along with disease histories which are personal and family-related.

The health states of people differ from person to person. And the health state of an individual changes over time [1]. Since we are living in a society of similar culture, customs, life styles, and diets, it is natural to believe that we develop similar health states associated with common health features and personal or family histories in adult or lifestyle diseases. This study explores a way of categorizing and inferencing over health states in terms of several common diseases, such as apoplexy, heart problem, and hypertension.

Health state changes over time. The change is believed to have a typical dynamics which, however, is subject to uncertainty due to unexplained variabilities and noisy or missing values. Being a temporal sequence, the health data adds another dimension of difficulty – time. Changes in health states look arbitrary and their speed vary depending on the age, life style, and heredity. By studying samples from a population, we can identify a set of clusters, each corresponding to a meaningful health state. Then by relating one state to another in the temporal order as observed in the data, we can capture the health dynamics that govern health state changes. We will focus on probabilistic interpretation to make predictions over health states.

### III. MODEL OF HEALTH STATE AND DYNAMICS

#### A. Health State

There are as many health states as there are people in the world. Notwithstanding we can categorize people according to their health in many ways. And if we limit our discussion to one particular aspect of a certain disease, the task can be made much simpler. Based on this reasoning we approach the problem of defining health states in two stages. First we extract a set of collections of health records according to personal disease histories. This research considers eight common disease types. But our discussion will center around one representative category for reasons of space. The states will differ between people with one disease history and those without it. Next, for each of the collections we developed a set of discrete health states, each in turn corresponding to a cluster of clinical records. These clusters are given by such records as estimated to have come from a particular state in the HMM for the disease history of interest. Unlike ordinary clustering methods which simply partition a set of static vectors, the clusters in the proposed method are not defined in isolation but in the temporal and causal context of states. Figure 1 shows a sample of two-mixture Gaussian emission densities for some variables in a ten-state HMM for people with and without apoplexy history respectively. Many Gaussians from different states have similar shapes and do appear overlapped in the feature space. But they can go separate ways in other dimensions and temporal context, thus distinguishing the health states.

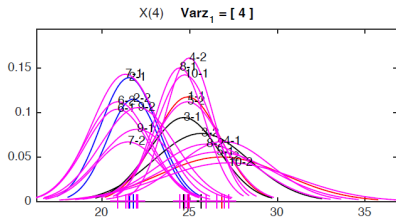


Figure 1. Gaussian states seen from BMI (body-mass index) values. Each Gaussian is labeled by a state-mixture identifier pair of an HMM for people without an apoplexy history.

Several overlapping bumps are obvious indicating the presence of distinct health states in different contexts.

#### B. Hidden Markov Model

HMM is a popular stochastic modeling tool for time series data. It generates an observation each time. It assigns a variable to each observation  $y_t$  which is a vector in our case. Given an observational record vector  $y_t = (y_{t,1}, y_{t,2}, \dots, y_{t,K}) = (y_{t,1}, y_{t,2}, \dots, y_{t,K})$  where  $K$  is the number of variable groups, we define the emission probability as a product of individual densities of the variable groups as follows

$$f_j(y_t) = \prod_{k=1}^K f_{jk}(y_{t,k}), \quad 1 \leq j \leq N, \quad 1 \leq t \leq T \quad (1)$$

where  $N$  is the number HMM states,  $T$  is the length of an input sequence. Each function  $f_{jk}()$  is a continuous or discrete distribution function depending on value of  $y_{t,k}$ . The graphical model for the HMM is shown in Figure 2. The model shows a set of conditional dependency relations among variables. Given a sequence of length  $T$ , the state variable  $X_t$  and observation vector  $y_t$  are duplicated  $T$  times left to right.

We create an HMM for each category of disease history and train it using the Baum-Welch algorithm. Since, however, there are many missing values, we have modified the algorithm that marginalizes out those variables. The number of states is shared by all models which are basically ergodic.

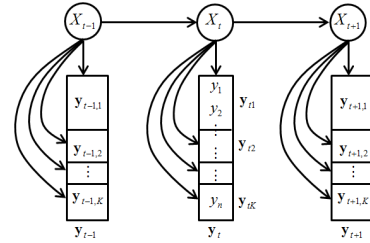


Figure 2. The graphical model of an HMM with multiple subvectors of conditionally independent observations.

#### C. Health State Dynamics

HMM describes the dynamic aspect of a state space using a first-order Markov chain as given by a stochastic matrix  $A = (a_{ij}) \in \mathcal{R}^{n \times n}$  which can be visualized as Figure 3. Each row of  $A$  defines a probability distribution and normally the diagonal elements dominate the probability distributions. A diagonal element  $a_{ij}$  represents the tendency of the Markov chain to stay in state  $i$ , be it a persistently healthy state or an unhealthy or even risky state.

Apart from the diagonal elements, there are a few elements with significant values in each row of the matrix, most of which being in the 1st, 2nd, 4th, 13th, 14th, 23rd, 26th, and 27th columns. We interpret it that they correspond to the popular destination nodes as is shown with many and strong incoming arrow heads in the right figure. They are the likely destination states of the future transitions, for the better or worse health states. They will serve the basis for predicting health developments.

If a Markov chain stays in a state  $i \in \{1, \dots, N\}$  for some period of time units before moving on to another state, we can assign the state to all the observations made there in. We collect such vectors for each state to obtain a cluster of health clinic records. We call it a *state-data cluster*  $C_i$ .

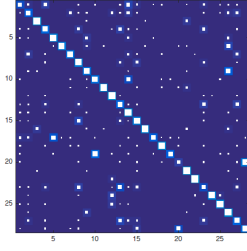


Figure 3. A sample stochastic matrix for a Markov chain with  $N = 28$  states for people with apoplexy history, given in a Hinton diagram where white squares represent transition probabilities greater than 0.01.

#### IV. PREDICTION

##### A. One-step Prediction

The transition matrix of a Markov chain provides a way of predicting the state one time unit ahead. Given a current state  $i$ , the prediction is best characterized by the  $i$ -th row of the matrix. We can make a prediction in two ways. First we can choose the state  $j^*$  of the highest probability. Then all the modes and or means of the output distributions for the  $K$  variable groups will serve the prediction:  $\hat{y}_{t+1}^{j^*}$ . An alternative way of prediction is applying the Bayesian rule to make a smoothed prediction based on the weighted average of the all the single state-based predictions

$$\hat{y}_{t+1} | i = \sum_{j=1}^N a_{ij} \hat{y}_{t+1}^{(j)}, \quad \forall i \in \{1, \dots, N\}, \quad t = 1, 2, \dots \quad (2)$$

For practical predictions as a public service, we instead focus on a subset of key variables that are of great concern for our healthy lives and consultation purpose. Such variables will include disease histories and risky symptoms related to adulthood and lifestyles.

##### B. Long-term Prediction

Although a Markov chain basically supports a unit step prediction, we can extend it beyond the immediate future by using the Chapman-Kolmogorov equation [15]:

$$A^{m+n} = A^m A^n \quad (3)$$

for some integers  $m, n \geq 0$  and set  $n = 1$  for our purpose. Then long term prediction is made pretty simple since all that is required is to prepare a sequence of powers of the matrix  $A^k$ ,  $k = 1, 2, \dots$ . We then conduct one-step prediction just by consulting the table of  $A^k$ 's.

In this paper we propose an alternative method of using state duration densities which are the derived statistics about outgoing transitions. All the state durations are known to follow approximately gamma distributions [16]. The proposed model is none other than the nonstationary HMM that was first introduced by Sin and Kim [17].

A prediction beyond the immediate future becomes increasingly unreliable as we extend it to farther out into the future. But if we provide more than one hypotheses as potential states, we can still make a meaningful prediction based on the HMM. Surprisingly enough, the predictive performance does not decrease exponentially but rather quickly saturates to a still noteworthy level, which has been confirmed from a set of experiments.

#### V. EXPERIMENTS AND ANALYSIS

##### A. Data Description/HMM Creation

This study has been conducted using the dataset obtained from National Health Insurance Sharing Service, Korea [18]. Each sample comprises a time sequence extending up to about twelve years. The dataset was divided multiple ways according to the presence of seven personal disease histories. For all the experiments to follow, we employed the set both for learning and testing, separately for each subset of samples. Table I shows the sizes of the classified sets.

Table 1. Data sets classified according to the seven personal disease histories. The figures are the number of sequences (or persons) in the collection. Some are truncated to 100,000.

Category names	Positive sequences	Negative sequences
Apoplexy	4,773	65,564
Heart disease	10,183	26,184
Hypertension	60,999	22,309
Diabetes	39,002	100,000
Hypertlipidemia	27,750	100,000
Pneumonia	10,492	100,000
Others including cancers	55,534	100,000

##### B. Model Decoding

The first test concerns viewing the trained HMMs in terms of basic modeling behaviors. We consider computing the most likely annotation given an observation sequence. We employ the Viterbi algorithm to get the best state sequences (shown in chain of circles) as shown in Figure 4 given input sequences (in filled rectangles). Each state sequence corresponds to an annotation of an observation sequence in terms of  $N = 10$  possible states. A state change implies a health state change. But a change in observation does not necessarily mean a change in states as shown in the figures. State changes occur due to complex interplay of many variables most of which are not shown here. Likewise, note that state changes can also occur when there is no change in the observation variables.

We then collect clinical records aligned to a state  $i$  into a set of vectors, which are termed the state cluster  $C_i$ . This will constitute a collective sample representation of a health state. Figure 5 provides a clustering for a pair of variables corresponding to body weight and height. The clusters are aligned linearly from lower left to top right. This justifies the use of the one-dimensional body-mass index. By looking at the individual variables, we can attach a health state label in common terms. This job is left to domain experts for clinical interpretation.

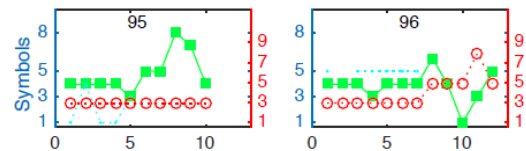


Figure 4. Two cases of model decoding given a sequence of values for a discrete variable shown in filled square markers

with the axis on the left. There are many more variables not shown here. Small circles represent the decoded states computed by the Viterbi algorithm.

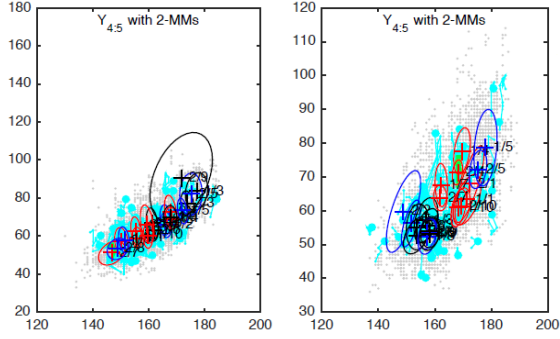


Figure 5. State-wise Gaussians modeling clusters of data points (dim dots). The two variables are height (cm, horizontal axis) and weight (kg, vertical axis). They are now combined to produce scalar BMI indices. (a) For people with apoplexy history, and (b) for those without one.

### C. State Change Behavior Model

A Markov chain changes states occasionally after a certain amount of time in a state. We note the distribution of durations in each state as a histogram is shown in Figure 6(a). The curves overlying the histograms represent gamma distributions for censored samples [19]. The gamma distribution has been considered useful as a parametric model of state durations in HMM states [17]. The light colored tops of the bars indicate the fraction of censored samples for various lengths of time durations in the corresponding states.

A Markov transition matrix describes a process dynamics with a set of stationary behaviors in the form of conditional distributions. In the standard HMM, then, the transition behavior in a state is fixed regardless of how long the process has maintained the state. In this research we explore the behavior by deriving a set of statistics over destination states of transitions as a function of duration length. Figure 6(b) shows the timed Markov transitions where each row of cells represents a distribution over destination states and for each  $(i, j)$  cell

$$dA_i(d, j) = \Pr(s_{t+1} = j \mid s_{t-d+1} = \dots = s_t = i \neq s_{t-d}, i \neq j), \quad (1)$$

$$0 \leq d \leq D_{max}, 1 \leq j \leq N,$$

denotes the probability of making a transition to state  $j$  from state  $i$  after a sojourn of  $d$  time units. Figure 6 shows two

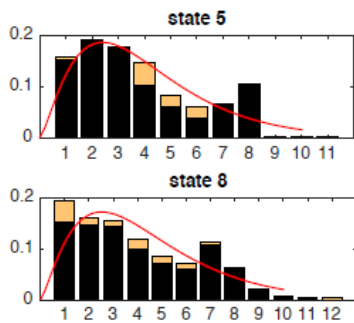


Figure 6. State duration statistics. State duration distributions of HMM for apoplexy with  $N = 10$  states derived from the

Viterbi decoding over the training set. Horizontal axis represents state durations in years.

examples. Being derived statistics from a Markov matrix, they actually add nothing new to the dynamics described in the matrix. But we will use both the Markov matrices and the derived duration statistics with similar effects for long-term prediction in health for comparison.

### D. Health State Prediction

This research aims to make predictions in health state as a service to the public. The first experiment concerns predicting health states one year ahead. The first-order Markov chain provides an answer to this question. For varied predictions, the duration-dependent transitions  $\{[dA_i]\}_{i=1,N}$  have been employed. Figure 7 shows sample sequences of single year predictions over time assuming that the future observations are not given yet. Although the prediction trajectories exhibit state changes occasionally out of time by one or two years, the prediction paths generally draw similar 'future' histories correctly. To be more specific, Figure 7(a) and (b) are prediction samples of persons with histories in apoplexy and hypertension respectively. Whereas Figure 7(c) and (d) show examples of prediction for persons without histories in apoplexy and heart problem respectively.

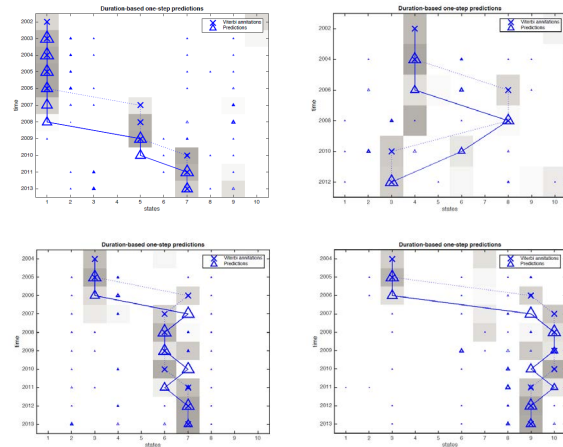


Figure 7. A sequence of single year predictions (triangle markers) for (a) person #14 with an apoplexy history and (b) person #6 with a hypertension history, (c) person #14 without apoplexy history, and (d) person #8 without heart problems yet. The 'x' markers denote the baseline Viterbi paths considered optimal for comparison.

HMM is a stochastic model whose complexity is determined by the number of states and the dimensionality of the observation vectors. The latter is not part of the adaptive parameters. But the former can be varied leading to the problem of model selection in machine learning. The next set of experiments involves detailed analysis of the performance in health state prediction as a function of the model size and the number of state hypotheses generated by the system.

Figure 8 shows a summary of the prediction performance in the group of people with apoplexy history. The two upper panels show prediction accuracies using the transition matrix and the derived duration/transition statistics. Since the latter has been derived from the former, their statistical behaviors are equivalent. According to the charts, the performance decreases with increasing number of states, which is naturally



due to the increased number of states to choose from; the more states there are in the space, the harder is the choice. But we can compensate the decline by increasing the number of hypotheses (bottom panels). This does make sense in prediction, or even sounds more convincing when such predictions are given with certain confidence measures. What is noteworthy is that the slopes of the curves are steep and approach 70% with three or four candidates even when the state space is  $N = 28$ . On the other hand Figures 9 and 10 present similar charts for different histories, heart problem and hypertension respectively.

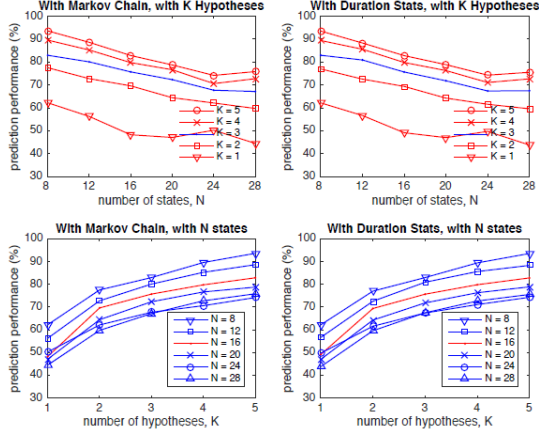


Figure 8. The performances of single year predictions for a range of values of  $K = 1, \dots, 5$ , and  $N = 8, \dots, 28$ , for people with apoplexy history.

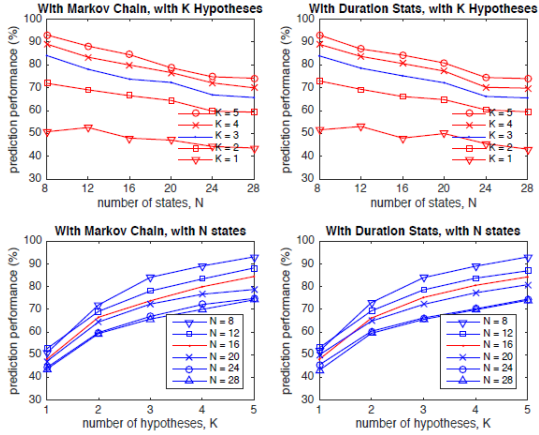


Figure 9. Single year prediction performance for people with heart problem histories.

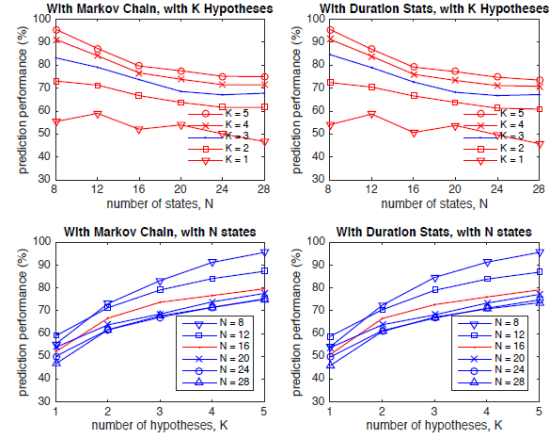


Figure 10. Single year prediction performance for people with hypertension histories.

The preceding tests show a general trend of performance decrease as the number of model states increases. Figure 11 provides a performance comparison of the three diseases over different number of states and hypotheses. The left panel shows the prediction using Markov chains and the right panel the prediction based on the derived statistics of differential transitions.

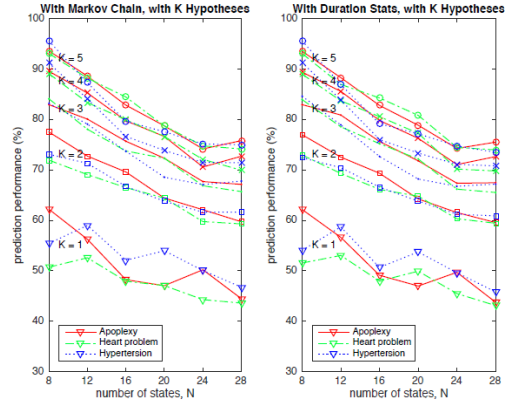


Figure 11. Single year prediction performance for comparison across disease histories. (a) Performance with Markov chains, and (b) those with derived differential transition statistics.

### E. Long-Term Prediction

Health clinic data is usually about annual checkups. Naturally we also find interest in the health state development based on the current state and/or any treatments and preventive activities. The standard HMM is basically a first-order Markov chain, which means that it captures the first-order correlation over time. This section tries to explore whether the model is effective for predictions beyond the first year.

The next set of tests involves a set of predictions extending up to five years which has been found to be statistically significant in the dataset. In addition for a more reliable prediction service, it is desirable to provide additional candidates for future states. Figure 12 shows the predictive performance for an HMM of apoplexy history with ten states up to six years with varying number of candidates. The front row of bars left to right represent the prediction performance up to six years, while the back-end row of bars correspond to the performance with five candidate states. The performance

starts at 59% in the first year and drops down to 37% in the 3rd to 6th years. But they all increase to 85% with five candidates regardless of the distance in years.

The prediction performance, however, is expected to decrease as the number of model states increases. Tables II and III show the prediction results regarding the development of three disease types. Table II lists the prediction rates of the three disease HMMs with a fixed number of states  $N = 16$ , while Table III highlights the effect of model sizes on the performance. What is noteworthy is the fact that the performance curves largely level off quickly in the second year, particularly as the number of models states increases.

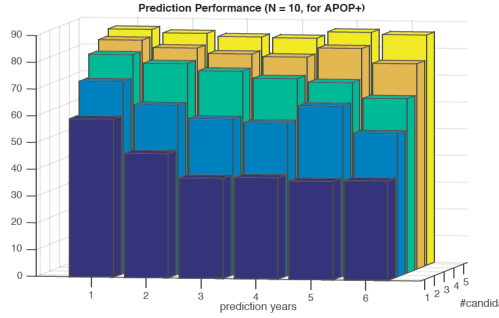


Figure 12. The performance of long-term predictions up to six years ahead, beyond which the result gets too unreliable to use. The front row of bars represents predictions up to six years (left to right) into the future with single best answers. The performance increases with increasing number of best candidates (rows of bars at the back).

Table II. Multi-year prediction performance (%) with HMMs with  $N = 16$  states and a varying number of hypotheses.

Prediction years		Number of candidates ( $K$ )				
		1	2	3	4	5
Apoplexy History	1	48.2356	69.6116	75.7045	79.7664	82.8637
	2	36.5430	60.3266	68.4587	73.2902	77.9177
	3	35.0696	53.5327	63.3316	70.9644	75.6060
	4	38.6994	50.6397	64.0725	68.3369	75.2665
	5	40.0612	49.5413	62.6911	66.9725	71.8654
Heart problem history	1	47.9438	66.5134	73.7676	79.9489	84.4700
	2	28.8264	55.0789	66.7811	75.1201	79.2725
	3	29.6238	46.2382	60.1881	71.5778	76.3323
	4	29.5405	43.3260	61.1597	70.2407	75.0547
	5	31.6294	45.0479	60.0639	67.7316	73.8019
Hypertension history	1	52.0948	66.7465	73.7132	76.6579	79.6505
	2	37.2951	59.5523	68.0958	71.9105	75.7566
	3	33.7177	54.3819	65.5443	70.8948	74.2159
	4	34.3966	54.1379	63.3621	70.0862	75.0862
	5	35.5172	53.2759	63.1034	71.5517	76.8966

Table III. Multi-year prediction with the apoplexy HMM with  $K = 1$  and varying number of model states.

Prediction years	Number of model states ( $N$ )					
	8	12	16	20	24	28
1	62.2239	56.2579	48.2356	47.0678	50.2158	44.4783
2	50.3232	43.0759	36.5430	30.7928	35.7605	29.2276
3	40.5782	36.2042	35.0696	28.8293	29.9639	23.2594
4	47.7054	42.7050	38.6994	30.2449	28.1450	27.7955
5	45.0920	44.6483	40.0612	33.6391	31.9018	29.4479

## VI. CONCLUSION

In this paper we first defined a set of health states in the context of dynamics manifested by the collection state observations using the tool of hidden Markov model (HMM). We also presented a method of predicting the future health states based on the Markov dynamics and state duration

statistics derived from the HMM and the training set. The proposed method returns a prediction in terms of probability distribution over the course of several years along with multiple candidates for a robust and flexible prediction. The prediction based on state duration is limited to state changes from the single current state. In order to overcome this limitation, we may have to resort to a sampling method in order to avoid the exponential complexity in long-term prediction. The current model is based on single state variable representing a hidden complex correlation among latent factors. If we decompose the state variables for a sparse and effective representation, we may expect a better modeling of the population health and its collective dynamics.

## REFERENCES

- [1] W. Ventres, S. Kooienga, N. Vuckovic, R. Marlin, P. Nygren, and V. Stewart, "Physicians, patients, and the electronic health record: an ethnographic analysis," *The Annals of Family Medicine*, vol. 4, n. 2, pp. 124-131, 2006.
- [2] L.R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257-286, 1989.
- [3] S.-T. Li, and Y.-C. Cheng, "A stochastic HMM-based forecasting model for fuzzy time series," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 40, no. 5, pp. 1255-1266, 2010.
- [4] T. Ribeiro, "Discrete dynamic models: A HMM approach to estimation and forecasting using panel survey data," UC Berkeley Working Paper, 2002.
- [5] R. Kawamoto, A. Nazir, A. Kameyama, T. Ichinomiya, K. Yamamoto, S. Tamura, et al., "Hidden Markov model for analyzing time-series health checkup data," *MedInfo*, pp. 491-495, 2013.
- [6] Z. Liu, Q. Li, X. Liu, and C. Mu, "A hybrid LSSVR/HMM-based prognostic approach," *Sensors*, vol. 13, no. 5, pp.5542-5560, 2013.
- [7] E. Choi, A. Schuetz, W. Stewart, and J. Sun, "Using recurrent neural network models for early detection of heart failure onset," *Journal of the American Medical Informatics Association*, p. ocw112, 2016.
- [8] T. W. Liao, "Clustering of time series data—a survey," *Pattern recognition*, vol. 38, no. 11, pp.1857-1874, 2005.
- [9] T. Oates, L. Firoiu, and P. R. Cohen, "Clustering time series with hidden Markov models and dynamic time warping," *Proceedings of the IJCAI-99 workshop on neural, symbolic and reinforcement learning methods for sequence learning*, pp. 17-21, 1999.
- [10] T. Wang, Trajectory similarity based prediction for remaining useful life estimation, PhD Thesis of University of Cincinnati, 2010.
- [11] O. F. Eker, Z. Skaf, F. Camci, and I. K. Jennions, "State-based Prognostics with State Duration Information of Cracks in Structures," *Procedia CIRP*, vol. 22, pp.122-126, 2014.
- [12] B. K. Sin, Y. W. Kim, and H. Choi, "HMM-based health state clustering and Prediction," *Proc. 5th Int. Conf. on Ubiqu. Computing App. and Wireless Sensor Netowrk (UCAWSN-16)*, 2016.
- [13] L.E. Baum, and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *The annals of mathematical statistics*, vol. 37, no. 6, pp. 1554-1563, 1966.
- [14] R. JA. Little, and D. B. Rubin, *Statistical analysis with missing data*, John Wiley & Sons, 2014.
- [15] A. Papoulis, and S. U. Pillai, *Probability, random variables, and stochastic processes*, Tata McGraw-Hill Education, 2002.
- [16] B. K. Sin, "Gamma CDF-based HMM state duration modeling," *J. Korea Institute of Info. Sci. and Eng.: SW and Applic*, vol. 40, no. 12, pp.757-763, 2013.
- [17] B. K. Sin, and J. H. Kim, "Nonstationary hidden Markov model," *Signal Processing*, vol. 46, no. 1, pp. 31-46, 1995.
- [18] <http://nhiss.nhis.or.kr/op/it/index.do>, National Health Insurance Sharing Service, Korea, accessed 25 December 2016.
- [19] H. L. Harter, and A. H. Moore, "Maximum-likelihood estimation of the parameters of gamma and Weibull populations from complete and from censored samples," *Technometrics*, vol. 7, no. 4, pp. 639-643, 1965.