Using Hidden Markov Models to Capture Temporal Aspects of Ultrasound Data in Prostate Cancer

Layan Nahlawi*, Farhad Imani†, Mena Gaed‡, Jose A. Gomez §, Madeleine Moussa§, Eli Gibson¶, Aaron Fenster‡¶, Aaron D. Ward¶, Purang Abolmaesumi†, Parvin Mousavi*and Hagit Shatkay*‖
*School of Computing, Queen's University, lnahlawi@cs.queensu.ca
†Dept. of Electrical and Computer Engineering, University of British Columbia
§Dept. of Pathology, Schulich School of Medicine and Dentistry, Western University Ontario
¶Dept. of Medical Biophysics, Western University Ontario, ‡Robarts Research Institute
‖Dept. of Computer and Information Sciences, University of Delaware

Abstract

Recent studies highlight temporal ultrasound data as highly promising in differentiating between malignant and benign tissues in prostate cancer patients. Since Hidden Markov Models can be used for capturing order and patterns in time varying signals, we employ them to model temporal aspects of ultrasound data that are typically not incorporated in existing models. By comparing order-preserving and order-altering models, we demonstrate that the order encoded in the series is necessary to model the variability in ultrasound data of prostate tissues. In future studies, we will investigate the influence of order on the differentiation between malignant and benign tissues.

1. Introduction

Prostate cancer is the second leading cause of death in male cancer patients, with a total of 27,540 estimated deaths in the United States for 2015 alone [1]. Core needle biopsy of the prostate under *Trans-Rectal UltraSound (TRUS)* guidance is used for definitive diagnosis of prostate cancer. TRUS guided biopsies are prone to both under-diagnosis and over-diagnosis [2]. In addition, TRUS guidance is not accurate enough for patient specific targeting in the prostate. The lack of accurate guidance is due to the low sensitivity and specificity of standard ultrasound imaging in differentiating between cancerous and normal regions in the prostate.

Recent research efforts are directed toward developing targeted biopsies that focus on regions with a high probability of being cancerous. These approaches augment biopsy procedures by providing tissue-specific information, from the analysis of the ultrasound signals or pre-operative Magnetic Resonance Imaging (MRI), in order to increase the probability of targeting cancerous regions [3]. Differentiation between cancerous and normal tissue is known as *tissue characterization* and is based on various types of ultrasound data analysis. In this work, we focus on analyzing temporal ultrasound radio frequency echo signals (hereinafter referred to as *RF time series*), which have been introduced by Moradi $et\ al.$ and have been since applied toward the detection of prostate cancer, breast cancer and ablated tissue [4]. An RF time series is a sequence of ultrasound frames captured from a specific location in the tissue, over a specific period of time, without intentionally moving the ultrasound transducer or the tissue. It relays the response of tissues to repetitive ultrasound irradiation. It also carries tissue-specific information that can be used to differentiate between cancer and normal tissues. Typically, tissue-specific information, is extracted from RF time series in the frequency domain (such as wavelet and mean central frequency) and used in a machine learning framework.

Since RF time series are time varying signals, we investigate here the utility of Hidden Markov Models (HMMs) to model the variability of these time series in cancerous vs normal prostate tissue. HMMs are stochastic models that capture temporal relations and order in time series. They are often used to model and analyze time series data under the simplifying assumption of the Markov property, namely, that the state at a certain time point depends only on the state at the preceding time point and conditionally independent of all other time points. HMMs can be learned in an unsupervised fashion, and can also be used in the supervised context to learn class-specific structures from labeled-data. Thus HMMs are widely used in machine learning. Rabiner $et\ al.$ demonstrated the use of HMMs primarily in the context of speech recognition [5]. HMMs are also often utilized to model biological data such as proteins and DNA sequences [6]. However, we are not aware of any previous work in which HMMs are utilized to model RF ultrasound data, which is the focus of the work presented here.

We hypothesize that the order of the values in the RF time series carries significant information needed to model the variability in the signal. To capture the

978-1-4673-6799-8/15/$31.00 ©2015 IEEE

temporal order of points in the RF signals and demonstrate its role in describing sequence variability, we use HMMs to model RF time series of prostate cancer patients based on their original temporal order as well as under various levels of order shuffling applied to the signals. We assess the difference between the order-preserving model and the order-altering model, by comparing the probability distribution induced by the HMM trained on the ordered RF time series to the distribution induced by the HMM trained on the shuffled series. For comparing models, we use a symmetric divergence measure based on the Kullback-Leiber (KL) divergence [7].

In the next section, we describe the data and its representation. We then present the proposed HMM and the comparison approach along with the results. Finally we provide conclusions and future work.

2. Data

The data was gathered (under informed consent and approval of IERB) via in-vivo and ex-vivo imaging of cancerous prostates, obtained from 10 patients who have undergone radical prostatectomy as part of their treatment (see [3] for details). After removal of the prostate, MRI images of the specimen were obtained [8]. The prostate was then sliced into 4.4 mm-thick sections; high-resolution imaging and annotation of the malignant areas was performed by a physician and confirmed by a genitourinary pathologist. 3D reconstruction of the tissue sections was then performed [9], and subsequently registered to in vivo ultrasound imaging as described in [3]. This process yields an overlay of the pathology assessments onto the utrasound images used as the ground truth.

Ultrasound imaging techniques do not provide accurate information about the location of very small objects, due to the scattering phenomenon: the echoes reflected from such small objects in soft tissues are scattered in all directions rather than solely in the direction back to the transducer. Thus, the annotation of ultrasound images is based on groups of RF values, not a single value, corresponding to areas that can be easily visualized by clinicians. Imani *et al.* proposed the use of RF values corresponding to areas, known as of Regions Of Interest (ROIs), of 1.7x1.7mm in size [3]. In this paper, we use the same ROI size.

The RF time series of each patient consists of 128 frames, where each frame corresponds to an image taken at a single time point and consists of 1276x64 intensity values. Each value reflects the echo intensity captured from a specific location in the screened area. Figure 1 shows an illustration of RF time series data

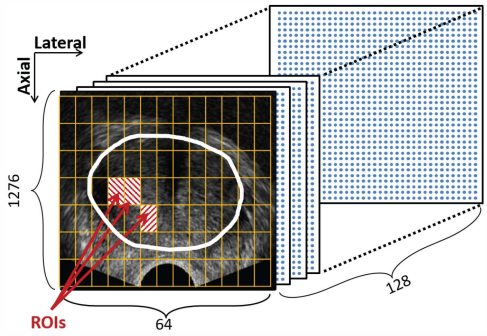

Figure 1: Ultrasound RF echo time series for one patient. It consists of 128 frames, each corresponding to a single point in time, where each frame comprises of 1276 pixels in the axial direction and 64 pixels in the lateral direction. The encircled area in the middle denotes the RF values corresponding to the prostate. The arrows are pointing at a group of selected regions of interest.

where the encircled area denotes the RF intensity values corresponding to the prostate. A grid of 1.7 x 1.7mm squares is overlaid on each frame. The grid divides each frame into ROIs of 44 RF values in the axial direction and 2 values in the lateral direction. The ROI grid is depicted in Figure 1. The arrows in Figure 1 point to selected ROIs. Data from the 10 subjects, with clearly labeled normal and cancerous ROIs, are used in the analysis described here. The available data consists of 625 selected ROIs of which 314 are cancerous and 311 are normal as selected by Imani *et al.* [3]. We use the mean intensities of all RF values in each selected ROI as the time domain feature modeled by the HMM. Due to the scattering phenomenon, the imaging technique does not accurately reflect single RF values, as a single RF value can be attributed to noise or may be generated by multiple scatterers. Therefore, averaging the information of 44x2 intensity values into a single mean value minimizes the impact of noise in the RF time series without significant loss of information, and helps reduce the dimensionality of the data.

Since the mean values are real numbers and our HMMs are based on discrete observations (see Section 3 for details,) we discretize the resulting time series values. The range of mean values is identified by finding the minimum and maximum values of ROI means in all patients. This range is then uniformly divided into Bn sub-ranges (bins). The ordinal number of each bin is used to represent the mean values falling in the sub-range values associated with the bin. We chose $Bn = 128$ based on the cost measure proposed by Shimazaki *et al.* [10]. To generate the shuffled RF time series, we randomly sample S pairs from the time points in each ROI and swap the values between each pair. We generated six shuffled versions of the RF time series with S equal to 10, 20, 30, 40, 50 and 64.

3. HMM for Order Assessment

We assume that the extent of the tissue response to ultrasound radiation at certain time-point, depends only on the response in the previous time point, and has no memory of prior time points. An HMM is defined as a pair of two stochastic processes where the first process consists of transitions among states that are unobservable and can only be estimated through the second process that generates a sequence of observed symbols [5]. In this work, we use a discrete HMM to model discretized RF time series signals. An HMM is formally defined using five elements: the set of N states $\{S_1, ...S_N\}$; the set of M observations $\{V_1, ...V_M\}$; the state transition probability distribution matrix, A; the observation symbol probability distribution matrix, B and the initial state distribution vector, Π [5].

In our proposed HMM, each time point in the RF signal corresponds to a state in the hidden model. Since RF time points have a left-to-right order, our HMMs are described as strictly left-to-right models. The only state transitions allowed are from the state at time point t to the state at time point $t+1$ for all time points $t = \{1, ...T-1\}$. The observations associated with the states are the bin numbers corresponding to echo values in the RF time series. We view each RF time series of a single ROI as a time varying stochastic random process that we model using HMMs. The length of our observation sequence is T =128 and we use the same number (128) as the number of states N; the set of possible observation symbols is the alphabet V = 1, 2, 3…128. The transition matrix A whose elements are a_{ij} is defined as:

$$a_{ij} = Pr(S_j \ at \ (t+1) | S_i \ at \ t) = 1, \forall j = i+1$$

where $i \in \{1, ...N-1\}, j \in \{2, ...N\}$ otherwise $a_{ij} = 0$,
while the observation matrix B whose elements are b_{jk} is defined as:

$$b_{jk} = Pr(V_k \ at \ t | S_j \ at \ t) = \frac{\#\ of\ V_k\ at\ t}{\#of\ symbols\ at\ t},$$

where $k \in \{1, ...M\}, j \in \{1, ...N\}$,
The initial vector Π is defined such that $\pi_1 = 1$ and $\pi_i = 0$ for all $i \neq 1$. We use the cancerous ROIs to train an HMM that corresponds to cancerous data (to which refer as the *cancerous HMM*) and the normal ROIs to train a *normal HMM*. In addition, we use the six shuffled versions of the RF time series to train another set of six cancerous and six normal HMMs.

The trained HMMs represent tissues response to prolonged ultrasound irradiation, over time. We ran multiple experiments to compare between ordered and shuffled models while changing the number of shuffled time points. We compare the HMMs to determine if the order of points in the RF time series carries information necessary to model the signals. The comparison of HMMs assesses how different the transition and emission probability distributions of HMMs trained with RF times series in their original order (order-preserving HMM referred to as ordered HMM) and the ones trained with shuffled RF time series (order-altering HMM referred to as shuffled HMM).

To compare HMMs, we use Monte-Carlo sampling to obtain pairs of simulated observation sequences, where in each pair, one sequence, Seq_i, is sampled from the ordered HMM, λ_i, and the other, Seq_j, is sampled from the shuffled HMM, λ_j. We then calculate the KL-divergence D_i of λ_i from λ_j and the KL-divergence D_j of λ_j from λ_i, as suggested by Juang $et\ al.$ [7]. The KL-divergence measure between two models λ_i and λ_j is calculated as:

$$D_i(\lambda_i, \lambda_j) = logP(Seq_i|\lambda_i) - logP(Seq_i|\lambda_j) \quad (1)$$

where Seq_i is a simulated sequence sampled from model λ_i, $logP(-|\lambda_i)$ and $logP(-|\lambda_j)$ are the log probabilities of a sequence given the models λ_i and λ_j. Since the divergence measure is non-negative and not symmetric, Juang $et\ al.$ propose the calculation of the average of the two KL-divergence values $D_i\ and\ D_j$, between two models λ_i and λ_j, as a symmetric KL-divergence measure $D_s = \frac{D_i + D_j}{2}$ [7].

If the order of points does not carry information about the RF time series and their variability, the ordered HMM and the shuffled HMM are expected to be almost equivalent and the KL-divergence between them is expected to be close to zero. When the compared HMMs are equivalent, the log probabilities of a sequence given each of the HMMs are identical and the divergence is zero. Otherwise if the HMMs are not equivalent, the divergence is expected to be a positive real number. As the difference between the models increases, the KL-divergence between them increases as well.

4. Results and Discussion

After training seven cancerous HMMs and seven normal HMMs (one ordered and six shuffled), we repeatedly generated 500 observation-sequences from each of the HMMs to overcome any bias introduced by the random shuffling and sampling of the observation sequences. Using an ordered sequence and a shuffled one we calculated the symmetric KL-divergence D_s. Each pair of the simulated sequences from the two compared models gives rise to a single D_s. We calculated 500 symmetric KL-divergence values between the ordered cancerous HMMs and each of the shuffled cancerous HMMs and similarly for each of the normal HMMs. We then averaged 500 KL-divergence values

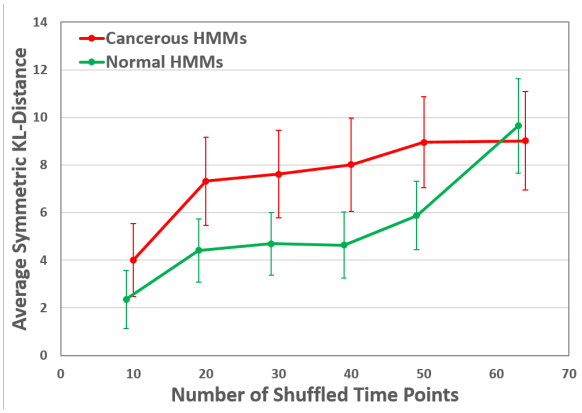

Figure 2: The average symmetric KL-divergence between order-preserving and order-altering HMMs as a function of the number of shuffled time points in the signal along with their standard deviations over 500 runs. The line with triangular markers shows the average KL-divergence for the HMMs of cancerous ROIs and the line with circular markers shows the KL-divergence for the HMMs of normal ROIs.

to obtain a mean estimate of the difference between ordered HMMs and each of the shuffled HMMs. As shown in Figure 2, we plot the mean KL-divergence values between the ordered and the shuffled HMMs as a function of the number of shuffled time points. The plot with triangular markers shows the results obtained when data from cancerous regions are used to train the HMMs (both the ordered and the shuffled models), whereas the plot with circular markers shows results obtained when the training data used comes from normal tissue regions (for both the ordered and the shuffled models). As shuffling the sequence order adds noise the signal, we anticipate an increase in the symmetric KL-divergence (between the ordered and shuffled HMMs) as the number of shuffled time points increases, which is indeed visible in the figure.

The increase in the symmetric KL-divergence between the ordered and shuffled HMMs as a function of the amount of shuffling demonstrates that the sequential order in the data carries significant information that can have value in modeling the response of prostate tissues to ultrasound irradiation in a probabilistic model. The graph does not show a difference in the trend (non-decreasing) between cancerous and non-cancerous models except for an insignificant decrease between point 30 and 40. However, the KL-divergence mean values of the cancerous models are larger than those of the normal models. The cancerous models appear more sensitive to random shuffling, until the point where the sequences are shuffled beyond recognition where more than 50% of the time points are substituted and nothing of the original order is retained.

5. Conclusion

We have proposed using a probabilistic modeling scheme based on HMMs for RF time series obtained from prostate cancer patients. Modeling the RF time series in the time domain using an HMM lets us capture the temporal order in the RF time series, which is not taken into account in previous studies. The increase in the value of mean symmetric KL-divergence, as the amount of shuffling introduced into the sequence increases, indicates that the order of points carries information necessary to model the variability in the responses of prostate tissues to ultrasound irradiation. In future work, we will explore the impact of incorporating temporal-order into the tissue-classification process, and its utility when using RF time series as a basis for characterizing normal vs cancerous tissues.

References

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2015," *CA: a cancer journal for clinicians*, vol. 65, no. 1, pp. 5–29, 2015.

[2] V. A. Moyer, "Screening for prostate cancer: US preventive services task force recommendation statement," *Annals of Internal Medicine*, vol. 157, no. 2, pp. 120–134, 2012.

[3] F. Imani, P. Abolmaesumi, E. Gibson, A. Khojaste, M. Gaed, M. Moussa, J. A.Gomez, C. Romagnoli, M. Leveridge, S. Chang, D. R. Siemens, A. Fenster, A. D. Ward, and P. Mousavi, "Computer-aided prostate cancer detection using ultrasound RF time series: In vivo feasibility study." *IEEE transactions on medical imaging*, 2015.

[4] M. Moradi, P. Mousavi, P. A. Isotalo, D. R. Siemens, E. E. Sauerbrei, and P. Abolmaesumi, "A new approach to analysis of RF ultrasound echo signals for tissue characterization: animal studies," in *Medical Imaging*. International Society for Optics and Photonics, 2007, pp. 65 130P–65 130P–10.

[5] L. 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.

[6] R. Hughey and A. Krogh, "Hidden markov models for sequence analysis: extension and analysis of the basic method," *Computer applications in the biosciences : CABIOS*, vol. 12, no. 2, pp. 95–107, Apr 1996.

[7] B.-H. F. Juang and L. R. Rabiner, "A probabilistic distance measure for hidden markov models," *AT&T technical journal*, vol. 64, no. 2, pp. 391–408, 1985.

[8] A. D. Ward, C. Crukley, C. A. McKenzie, J. Montreuil, E. Gibson, C. Romagnoli, J. A. Gomez, M. Moussa, J. Chin, G. Bauman *et al.*, "Prostate: Registration of digital histopathologic images to in vivo MR images acquired by using endorectal receive coil," *Radiology*, vol. 263, no. 3, pp. 856–864, 2012.

[9] E. Gibson, C. Crukley, M. Gaed, J. A. Gómez, M. Moussa, J. L. Chin, G. S. Bauman, A. Fenster, and A. D. Ward, "Registration of prostate histology images to ex vivo MR images via strand-shaped fiducials," *Journal of Magnetic Resonance Imaging*, vol. 36, no. 6, pp. 1402–1412, 2012.

[10] H. Shimazaki and S. Shinomoto, "A method for selecting the bin size of a time histogram," *Neural computation*, vol. 19, no. 6, pp. 1503–1527, 2007.