Data Mining of the E-Pelvis Simulator Database: A Quest for a Generalized Algorithm for Objectively Assessing Medical Skill

Thomas MACKEL [1], Jacob ROSEN [1,2], Ph.D., Carla PUGH [3], M.D., Ph.D.

[1] *Department of Electrical Engineering,*
[2] *Department of Surgery,*
University of Washington, Seattle, WA, USA
[3] *Department of Surgery,*
Northwestern University, Chicago, IL, USA

E-mail:{tmackel ,rosen}@u.washington.edu drpugh@northwestern.edu
Biorobotics Lab: http://brl.ee.washington.edu

Abstract: Inherent difficulties in evaluating clinical competence of physicians has lead to the widespread use of subjective skill assessment techniques. Inspired by an analogy between medical procedure and spoken language, proven modeling methods in the field of speech recognition were adapted for use as objective skill assessment techniques. A generalized methodology using Markov Models (MM) was developed. The database under study was collected with the E-Pelvis physical simulator. The simulator incorporates an array of five contact force sensors located in key anatomical landmarks. Two 32-state fully connected MMs are used, one for each skill level. Each state in the model corresponds to one of the possible combinations of the 5 active contact force sensors distributed in the simulator. Statistical distances measured between models representing subjects with different skill levels are sensitive enough to provide an objective measure of medical skill level. The method was tested with 41 expert subjects and 41 novice subjects in addition to the 30 subjects used for training the MM. Of the 82 subjects, 76 were classified correctly (92%). Moreover, unique state transitions as well as force magnitudes for corresponding states (expert/novice) were found to be skill dependent. Given the white box nature of the model, analyzing the MMs provides insight into the examination process performed. This methodology is independent of the modality under study. It was previously used to assess surgical skill in a minimally invasive surgical setup using the Blue DRAGON, and it is currently applied to data collected using the E-Pelvis.

1. Introduction

Inherent difficulties in evaluating clinical competence of physicians has lead to the widespread use of subjective skill assessment techniques. Subjective evaluation techniques lead to inconsistent evaluation by different examiners. Inspired by an analogy between medical procedure and spoken language, proven modeling methods in the field of speech recognition were adapted for use as objective skill assessment techniques. A metric that represents the skill level of a subject was determined by analyzing the subject's performance with respect to the performance of subjects of known skill levels. Previous studies applied the Markov modeling (MM) approach to skill evaluation of Minimally Invasive Surgery [1-6]. Using an approach that is independent of the modality used by the physician, the aim of the current study was to utilize the MM approach in developing a methodology for objectively assessing clinical skills during a pelvic exam using data acquired with the E-Pelvis simulator [7-9].

Medicine Meets Virtual Reality 14, Long Beach CA, January, 2006
J.D. Weswood et. Al. (Eds.), IOS Press, 2006

2. Methods

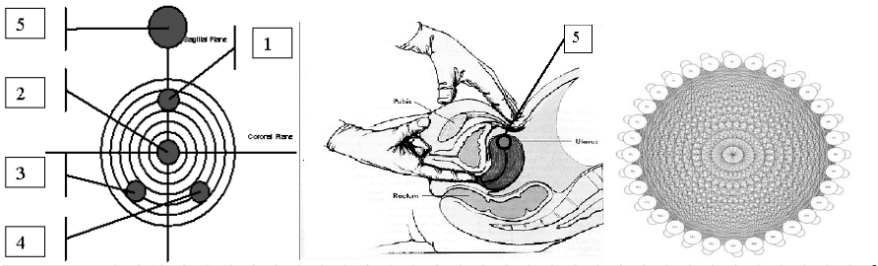
2.1. The E-Pelvis Simulator and the Acquired Database

The E-Pelvis is a physical simulator, shown in Figure 1, which consists of a partial mannequin (umbilicus to mid thigh) constructed in the likeness of an adult human female [7-9]. The simulator sampled data at 30 Hz from 5 pressure sensors located on key anatomical structures while the subjects performed pelvic examinations (Figure 2). The 41 expert subjects were selected from 362 professional examiners. The 41 novice subjects were selected from a group of 82 students. A different set of subjects, 15 experts and 15 novices, were selected to train the Markov models.

Figure 1. The Complete E-Pelvis Simulator, a Simulated Pelvic Exam, and the Graphical User Interface.

2.2. Data Analysis

An analogy between the human spoken language and minimally invasive surgery tasks [1-6] was extended to pelvic examination tasks. Based on this analogy, the primary elements, 'words', were the actual states of the MMs. Different 'pronunciations' of each state were observed in the pressure data taken from the E-Pelvis simulator. Data characterizing the performance of two categories of medical examiners, expert and novice, were analyzed using two 32-state fully connected MMs (Figure 2). Within each model certain sequences of state transitions, known as Markov chains, are more probable than others.



State		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
Sensor	1																	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	2									x	x	x	x	x	x	x	x									x	x	x	x	x	x	x	x
	3					x	x	x	x					x	x	x	x					x	x	x	x					x	x	x	x
"x" marks an	4			x	x			x	x			x	x			x	x			x	x				x	x			x	x		x	x
active sensor	5		x		x		x		x		x		x		x		x		x		x		x			x		x			x		x

Figure 2. **Top**: Sensor Locations and a 32-State Fully Interconnected Markov Model State Diagram.
Bottom: Active/Inactive Sensor Combinations

Medicine Meets Virtual Reality 14, Long Beach CA, January, 2006
J.D. Weswood et. Al. (Eds.), IOS Press, 2006

Each subject's performance was represented as a 5-dimensional vector of *N* data points measured from the 5 simulator sensors. A sensor was considered active if the value of the data exceeded a chosen threshold value, and it was considered inactive otherwise. Each state in the MM corresponded to one of the possible combinations (Figure 2) of active sensors. Many states were more commonly used than others, resulting in an uneven distribution of the data points between states.

Each state can be characterized by a mean vector and covariance matrix ([B] matrix – continuous version observation). The frequency transition matrix ([A] matrix) defines the frequency in which transitions occur between the various states. The data points representative of each state were used to compute the 1x5 mean vector, μ, and 5x5 covariance matrix, Σ. To calculate the elements of the [A] matrix the number of state transitions in the training subjects' data were tallied and then normalized to 1.

Bayes' Decision Rule was used to classify an unknown subject as either an expert or novice. If there are two classes, A and B, this rule states to choose class A if P(A) > P(B), choose class B otherwise. Define observation vector O as a sequence of data points x[n]. Let P(A) be the probability of a sequence of data points arising from an expert model, $P(O|\lambda_{ES})$, and P(B) the probability of a sequence of data points arising from a novice model, $P(O|\lambda_{NS})$.

The probability that model λ would generate an observation vector O, $P(O|\lambda)$, is the product of probabilities that each data point x was produced by model λ, $P(x|\lambda)$.

$$P(O \mid \lambda) = \prod_{i=1}^{N} P(\mathbf{x}[i] \mid \lambda) \tag{1}$$

$P(x|\lambda)$ can be defined as the product of the membership probability ('B' matrix) P_M and the transition probability P_T ('A' matrix). The membership probability is defined using the total probability rule.

$$P_M(\mathbf{x} \mid \lambda) = \frac{p(\lambda) L(\mathbf{x} \mid \lambda)}{\sum_{j=1}^{M} p(\lambda_j) L(\mathbf{x} \mid \lambda_j)} \tag{2}$$

where $p(\lambda)$ represents the *a priori* probabilities of each model, $L(x|\lambda)$ is the likelihood of data point x belonging to model λ, and M is the total number of models. Eq. (2) is simplified by assuming identical *a priori* probabilities for each model.

$$P_M(\mathbf{x} \mid \lambda) = \frac{L(\mathbf{x} \mid \lambda)}{\sum_{j=1}^{M} L(\mathbf{x} \mid \lambda_j)} \tag{3}$$

The likelihood is modeled by the multivariate normal probability density function.

$$L(\mathbf{x} \mid \lambda) = \frac{1}{(2\pi)^{p/2} \sqrt{\|\Sigma\|}} e^{\frac{-(\mathbf{x}-\mu)^T \Sigma^{-1} (\mathbf{x}-\mu)}{2}} \tag{4}$$

where p is the number of dimensions of data, Σ is the covariance matrix of the model, and μ the mean vector of the model.

Medicine Meets Virtual Reality 14, Long Beach CA, January, 2006
J.D. Weswood et. Al. (Eds.), IOS Press, 2006

The transition probability is given by the transition matrix for the model, A_λ. This probability takes into account the probability of transitioning from the previous state S_1 to the current state S_2, and is found by directly indexing the model's transition matrix. The first data point is a special case, as the previous state is unknown, and P_T is assigned a value of 1 for this case.

$$P_T(\lambda) = \mathbf{A}_\lambda[S_1, S_2] \tag{5}$$

Given a sequence of data associated with a specific subject, the above method can be used to estimate the probability that the MM of a class generated the sequence. The subject can be classified as a member of the class whose model results in the highest probability.

More data is collected during slower examinations. This penalizes both models by increasing the length of the observation vector O, hence increasing the number of factors used to compute $P(O|\lambda)$. As a result, the $P(O|\lambda)$ of one subject cannot be directly compared to the $P(O|\lambda)$ of another. Subjects' behavior relative to one another cannot be measured without the use of a common benchmark. One method uses a third MM trained from the subject's own data samples, $P(O|\lambda_{OS})$, which is compared to the novice model and the expert model. Two statistical factors can be defined as:

$$NSF = \log(P(O \mid \lambda_{OS})) / \log(P(O \mid \lambda_{NS})) \tag{6}$$

$$ESF = \log(P(O \mid \lambda_{OS})) / \log(P(O \mid \lambda_{ES})) \tag{7}$$

where O is an observation vector representing the subject's performance, λ_{OS} is a subject model trained by the data O, and λ_{ES} and λ_{NS} are models trained by data from experts and novices respectively.

There are two ways of finding $P(O|\lambda_{OS})$: a competing method and a non-competing method. In the competing method, $P_M(O|\lambda_{OS}) + P_M(O|\lambda_{ES}) + P_M(O|\lambda_{NS}) = 1$ for each observation point. An increase in $P_M(O|\lambda_{OS})$ is accompanied by a decrease in $P_M(O|\lambda_{ES})$ and $P_M(O|\lambda_{NS})$.

In the non-competing method, $P_M(O|\lambda_{OS}) = P_M(O|\lambda_{ES}) + P_M(O|\lambda_{NS}) = 1$ for each observation point. Only the subject transition matrix influences the value of $P(O|\lambda_{OS})$. This method prevents the subject model itself from influencing the classification results.

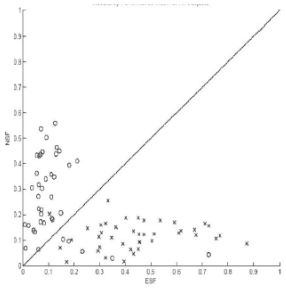
Given the 'white box' nature of the MM, analyzing the models provides insight into the process in which the pelvic exam is performed. Observing the most probable transitions within each model's transition matrix can identify Markov chains. The most-probable transitions are known as top-level chains. Some top-level chains are distinct to each class, and do not occur even as a subset of a chain of either class.

3. Results

Using the MMs to compute Bayesian classifier probabilities, 40 of the 41 expert subjects (97.6%), and 36 of the 41 novice subjects (87.8%) were correctly classified using this method. Overall, 76 of the 82 subjects tested (92.7%) were correctly classified. These subjects were not used in the training of the MMs. The skill factors computed using the non-competing method is shown in Figure 3.

Medicine Meets Virtual Reality 14, Long Beach CA, January, 2006
J.D. Weswood et. Al. (Eds.), IOS Press, 2006

Analysis of Variance (ANOVA) was performed to compare the data from each novice state with the corresponding expert state. Direct comparison of a set of novice data to a set of expert data showed interaction effects significant at the α=0.000001 level for all states except 18, 22, and 27, which were significant at the α=0.03 level. Expert data compared with another set of expert data showed interaction effects significant at the α=0.009 level or higher for all states, and greater than α=0.1 for most states. Novice data compared with novice data showed interaction effects significant at the α=0.00001 level for all states, and greater than α=0.1 for most states.



Expert Chains (11)	*Novice Chains (9)*
11-3-1	10-9-1
18-17-1	11-9-1
22-6-5-1	12-4-3-1
23-19-3-1	23-21-17-1
23-19-17-1	14-13-9-1
22-21-17-1	23-7-3-1
13-29-25-9-1	15-7-3-1
11-27-25-9-1	16-8-7-3-1
28-26-25-9-1	6-8-7-3-1
10-2-4-3-1	
18-2-4-3-1	

Figure 3. **Left**: Skill Factor Plot with Decision Boundary **Right**: Distinct Top-Level Markov Chains

Observation of the internal structure of the MM indicates that each skill level can be characterized by unique state transitions (Markov chains). 20 top-level chains distinct to only one model are shown in Figure 3.

4. Discussion

The use of MMs for data analysis has been successfully applied to speech recognition. Extension of this concept to objective medical skill assessment has lead to a successful Bayesian dichotomous classification method. Subjects classified in this manner can be compared to one another using their performance indices. The distinct chains identified in the models may help reduce the number of states in the models. The strength of this methodology is that it is independent of the modality under study. It was previously used to assess surgical skill in a minimally invasive surgical setup using the Blue DRAGON, and it is currently applied to data collected using the E-Pelvis as a physical simulator. Similarly, the same methodology can be incorporated into a surgical robot as a supervisory controller that could detect potentially dangerous mistakes by a human or computer operator.

References

[1] Rosen J, Hannaford B, Richards CG, Sinanan MN. Markov Modeling of Minimally Invasive Surgery Based on Tool/Tissue Interaction and Force/Torque Signatures for Evaluating Surgical Skills. IEEE Transactions on Biomedical Engineering, Vol. 48, No. 5, May 2001.

[2] Rosen J., M. Solazzo, B. Hannaford, M. Sinanan, Objective Evaluation of Laparoscopic Skills Based on Haptic Information and Tool/Tissue Interactions, Computer Aided Surgery, Volume 7, Issue 1, pp. 49-61 July 2002

[3] Rosen J., J. D. Brown, M. Barreca, L. Chang, B. Hannaford, M. Sinanan, The Blue DRAGON - A System for Monitoring the Kinematics and the Dynamics of Endoscopic Tools in Minimally Invasive Surgery for Objective Laparoscopic Skill Assessment, Studies in Health Technology and Informatics - Medicine Meets Virtual Reality, Vol. 85, pp.412-418, IOS Press, January 2002.

[4] Rosen J., J. D. Brown, L. Chang, M. Barreca, M. Sinanan, B. Hannaford, The Blue DRAGON - A System for Measuring the Kinematics and the Dynamics of Minimally Invasive Surgical Tools In–Vivo,

Proceedings of the 2002 IEEE International Conference on Robotics & Automation, Washington DC, USA, May 11-15, 2002.

[5] Rosen J., L. Chang, J. D. Brown, B. Hannaford, M. Sinanan, R. Satava, Minimally Invasive Surgery Task Decomposition - Etymology of Endoscopic Suturing, Studies in Health Technology and Informatics - Medicine Meets Virtual Reality, vol. 94, pp. 295-301, IOS Press, January 2003

[6] Kowalewski T.M., J. Rosen, L. Chang, M. Sinanan, B. Hannaford, Optimization of a Vector Quantization Codebook for Objective Evaluation of Surgical Skill, Studies in Health Technology and Informatics - Medicine Meets Virtual Reality, vol. 98, pp. 174-179, IOS Press, January 2004

[7] Pugh CM, Srivastava S, Shavelson R, Walker D, Cotner T, Scarloss B, et al. The effect of simulator use on learning and self-assessment: the case of Stanford University's E-Pelvis simulator. Stud Health Technol Inform 2001;81: 396-400

[8] Pugh CM, Rosen J., Qualitative and quantitative analysis of pressure sensor data acquired by the E-Pelvis simulator during simulated pelvic examinations, Stud Health Technol Inform. 2002;85:376-9.

[9] Pugh CM, Youngblood P. Free in PMC, Development and validation of assessment measures for a newly developed physical examination simulator. J Am Med Inform Assoc. 2002 Sep-Oct;9(5):448-60