Review of Sparse Hidden Markov Models for Surgical Gesture Classification and Skill Evaluation

Michael Norris

April 19, 2016

Sparse Hidden Markov Models for Surgical Gesture Classification and Skill Evaluation
Lingling Tao, Ehsan Elhamifar, Sanjeev Khudanpur, Gregory D. Hager, Ren Vidal

1 Introduction

This paper proposes a new model for classifying surgical gestures using sparse dictionary learning and hidden Markov models.

A surgeme is a surgical gesture. Examples include inserting a needle, grabbing a needle, or positioning a needle. A surgeme is an atomic unit that cannot be divided. For instance, it doesn't make sense to divide inserting a needle into multiple steps.

A motif is made up of one or more surgemes. Motifs are like the grammar of spoken language: they constrain words to certain patterns of phonemes, the atomic sounds that make up words. In language, phonemes only make sense in the context of words, and likewise a motif is a higher level description of the purpose of surgical gestures.

One or more surgemes become a maneuver, and one or more maneuvers make up a procedure.

The concepts in this paper are novel because they use concepts from natural language processing and speech recognition to uncover gestures in the way that we could try to understand a new language.

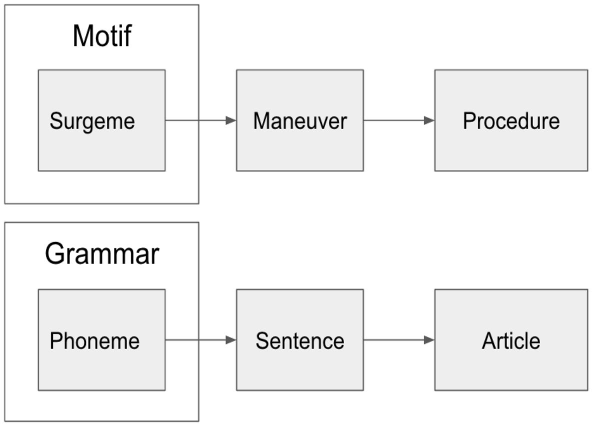

Figure 1. The relationship between language and surgical gestures.

The main problem with trying to model surgery like a language is that we don't know the words (maneuvers) or grammar (motif). We can, however, model the transitions between surgical gestures using a hidden Markov model to better understand the structure of gestures, and to develop a classifier.

"Given a surgery trial $\{y_t \in \mathbb{R}^D\}_{t=1}^T$, the goal of gesture classification is to assign a surgeme label $s_t \in \{1, \ldots, S\}$ to each frame y_t" [1, p. 2-3]
Similarly, the classification of a skill level involves assigning a skill level $z \in \{1, \ldots, L\}$ to the whole surgery $\{y_t \in \mathbb{R}^D\}_{t=1}^T$ [1, p. 2-3].

The surgeme label s_t is "hidden," and it has a transition probability of
$q_{s',s} = p(s_t = s | s_{t-1} = s')$ [1, p. 3].

In a sparse HMM, y_t is a sparse linear combination of elements from a dictionary of "motion words" [**?**, p. 3].

An observation at time t is:
$y_t = D_{s_t} x_t + e_t$, where $D_{s_t} \in \mathbb{R}^{D \times N}$
D is an over-complete dictionary (D ¡ N) at time t.

$x_t \in \mathbb{R}^N$ is a sparse latent variable
e_t is independent Gaussian Noise $N(0, \sigma_{s_t}^2 I)$

The distribution of y_t given the latent variable is:
$p(y_t | s_t = s, x_t = x) = N(D_s x, \sigma_s^2 I)$
To make x a sparse latent variable, use Laplacian prior on the distribution for x:

$p(x_t|s_t = s) \equiv \frac{\lambda_s}{2}^N$

Given N trials $\{y_{1:T_j}\}_{j=1}^J$ and surgeme labels $\{s_{1:T_j}^j\}_{j=1}^J$, we want to learn the sparse hidden Markov model.
The sparse hidden Markov model is the transition probabilities $Q = \{q_{s,s'}\}_{s,s'=1,...,S}$ and the parameters for each surgeme model:
$\Theta_s = (D_s, \sigma_s^2, \lambda_s)$, for $s = 1, \ldots, ..., S$

The transition probabilities are the probabilities that a gesture s will become gesture s' in the next frame. Since this is a Markov process, we don't have to worry about any other frame except the current one affecting the future.

Our model will learn which labels are most likely for the next gesture given the current one. We should expect that for gestures that are of a few seconds in duration, the transition probability for s_t to s_{t+1}, where $s_t = s_{t+1}$ should be highest regardless of the gesture.

The parameters for each surgeme model are the dictionary that represents it, the standard deviation σ for the noise, and the λ parameter of the Laplacian.

1.1 KSVD

The paper proposes that we use KSVD to learn the parameters for

1.2 Surgeme Classification

Given a trial $\{y_t\}_{t=1}^T$ and the S-HMM parameters $q_{s,s'}$ and Θ_s for $s, s' = 1, \ldots, S$, we want to infer the sequence of surgeme labels $\{s_t\}_{t=1}^T$ [1, p. 6]
In standard HMM's Viterbi can be used, but since we made the hidden states have a Laplacian, the paper proposes to use Basis Pursuit or Orthogonal Matching Pursuit [1, p. 6]. Both of these algorithms are finding a sparse representation for x in the already-known linear system D that maximizes the chance that we will predict the correct gesture.
Essentially, the algorithm needs to solve for $\hat{x}$ that maximizes $p(y_t|x, s)p(x|s)$, or:
$\hat{x} = \text{argmin}_x \lambda |x|_1 + \frac{1}{2\sigma_s^2}|y_t - D_s x|^2$

Since KSVD uses OMP, the authors use OMP.[1, p. 6]

1.3 Skill Classification

Skill classification is done by learning a Sparse HMM for each of three skill levels, expert, intermediate, and novice, and then classifying the trial by calculating the probability of the observed states, latent variables, and surgeme labels given each expertise model. The expertise model which gives the maximum score is the surgeon's skill level. [1, p. 6, equation 10]

2 Experiments

The data used from experiments came from a 78-feature motion dataset from a Da Vinci system. 8 surgeons of 3 different skill levels performed 3-5 trials each of 3 tasks.

2.1 Setup 1

Leave out one trial from each user for testing

2.2 Setup 2

Leave one user out for testing

2.3 Tasks

- Suturing - 39 trials
- Needle Passing - 26 trials
- Knot Tying - 36 trials

2.4 Ground Truth

Truth was created from video sequences of each task by manually labeling frames with surgeme labels.

3 Results

3.1 Results

The KSVD-HMM performed the best in Setup 1 (81.1%)for the suturing task. It performed well in Setup 2 as well, (67.8%). I hypothesize that Setup 2 has a lower classification rate because the left-out surgeon's gestures were most likely different than the other's. In Setup 1, we train with every surgeon's gestures. This shows that a single surgeon will perform tasks similarly between trials, but surgeons will not necessarily perform similarly compared with eah other.
For the Needle Passing, The Switched Linear Dynamical Systems approach had the highest classification rate.

For Knot Tying, KSVD-HMM was second highest, but was with .2% of the best classified algorithm, Factor Analyzed HMM

References

[1] Lingling Tao, Ehsan Elhamifar, Sanjeev Khudanpur, Gregory D. Hager, Ren Vidal *Sparse Hidden Markov Models for Surgical Gesture Classification and Skill Evaluation*, Third International Conference, IPCAI 2012, Pisa, Italy, June 27, 2012., Proceedings Information Processing in Computer-Assisted Interventions, Volume 7330 of the series Lecture Notes in Computer Science pp 167-177, 2011.