QUALITY CONTROL OF VOICE RECORDINGS IN REMOTE PARKINSON'S DISEASE MONITORING USING THE INFINITE HIDDEN MARKOV MODEL

AALBORG UNIVERSITY DENMARK

Aston University Birmingham

Amir Hossein Poorjam [1], Yordan P. Raykov [2], Reham Badawy [2], Jesper Rindom Jensen [1], Mads Græsbøll Christensen [1], Max A. Little [2,3]

[1] Audio Analysis Lab, CREATE, Aalborg University, DK [2] Engineering and Applied Science, Aston University, Birmingham, UK [3] Media Lab, MIT, Cambridge, Massachusetts, USA

[1] {ahp, jrj, mgc}@create.aau.dk, [2] yordan.raykov@gmail.com [2] rehambadawy@hotmail.com [2] max.little@aston.ac.uk

Introduction

- The performance of voice-based systems for remote monitoring of Parkinson's disease is highly dependent on the degree of adherence of recordings to the test protocols, which probe for specific symptoms.
- In this paper, we propose a method to automatically identify the segments of signals that adhere to/violate the test protocol:
  - **Fit an infinite HMM to the frames of the signal in the cepstral domain.**
    - Signal is split into variable duration segments.
  - **Apply a multinomial naïve Bayes classifier to the state indicator**
    - Identify segments that adhere to the test protocol.

Segmentation with infinite HMM

- **In HMM the likelihood of the observation x_t is modeled as:**

$$P(\boldsymbol{x}_t|s_{t-1}=i,\boldsymbol{\Theta})=\sum_{k=1}^{K}\pi_{i,k}P(\boldsymbol{x}_t|\boldsymbol{\theta}_k) \qquad (1)$$

To relax the assumption of a fixed K in (1):

$$\boldsymbol{\beta}\sim\text{Dirichlet}\left(\frac{\gamma}{K},\cdots,\frac{\gamma}{K}\right) \qquad \boldsymbol{\pi}_k\sim\text{Dirichlet}(\alpha\boldsymbol{\beta}) \qquad (2)$$

As $K\rightarrow\infty$, the hierarchical prior becomes a hierarchical Dirichlet process:

$$G_0\sim\text{DP}(\gamma,H) \qquad G_k\sim\text{DP}(\alpha,G_0) \qquad (3)$$

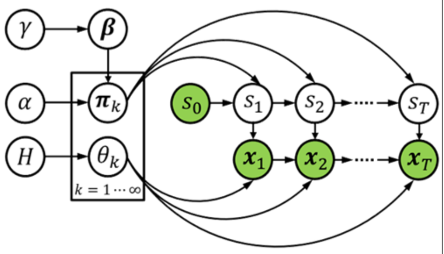
Under stick-breaking representation:

$$G_0=\sum_{j=1}^{\infty}\beta_j\delta_{\boldsymbol{\theta}_j} \qquad G_k=\sum_{j=1}^{\infty}\pi_{kj}\delta_{\boldsymbol{\theta}_j} \qquad (k=1,2,\dots,\infty) \qquad (4)$$

- **Infinite HMM model:**

$$\begin{aligned}\boldsymbol{\beta} &\sim \text{GEM}(\gamma)\\ \boldsymbol{\pi}_k &\sim \text{DP}(\alpha,\boldsymbol{\beta})\\ \boldsymbol{\theta}_k &\sim H\\ s_0 &= 1\\ s_t|s_{t-1} &\sim \boldsymbol{\pi}_{s_{t-1}}\\ \boldsymbol{x}_t|s_t &\sim f(\boldsymbol{\theta}_{s_t})\end{aligned}$$

For: $k=1,2,\dots,\infty$
$t=1,2,\dots,T$



The graphical model for the infinite HMM

Inference

- Direct assignment Gibbs sampler for inferring the posterior over sequence of hidden states.
  - **Resample $s_{1:T}$:** take out one s_t at a time and resample it from the posterior

$$P(s_t|s_{t'},\boldsymbol{x}_{1:T},\alpha,\gamma,\boldsymbol{\beta},H,f)\propto P(\boldsymbol{x}_t|s_t,s_{t'},\boldsymbol{x}_{t'},H,f)P(s_t|s_{t'},\alpha,\gamma,\beta) \qquad (5)$$

$$P(\boldsymbol{x}_t|s_t,s_{t'},\boldsymbol{x}_{t'},H,f)=\int P(\boldsymbol{x}_t|\boldsymbol{\theta}_{s_t})P(\boldsymbol{\Theta}|s_{t'},\boldsymbol{x}_{t'},H)d\boldsymbol{\Theta} \qquad (6)$$

$$P(s_t|s_{t'},\alpha,\beta)\propto\begin{cases}\left(n_{s_{t-1},k}+\alpha\beta_k\right)\dfrac{n_{k,s_{t+1}}+\alpha\beta_{s_{t+1}}}{n_{k,\cdot}+\alpha} & \text{for } k\le K,\ s_{t-1}\neq k\\ \left(n_{s_{t-1},k}+\alpha\beta_k\right)\dfrac{1+n_{k,s_{t+1}}+\alpha\beta_{s_{t+1}}}{1+n_{k,\cdot}+\alpha} & \text{for } s_{t-1}=s_{t+1}\neq k\\ \left(n_{s_{t-1},k}+\alpha\beta_k\right)\dfrac{n_{k,s_{t+1}}+\alpha\beta_{s_{t+1}}}{1+n_{k,\cdot}+\alpha} & \text{for } s_{t-1}=k\neq s_{t+1}\\ \alpha\beta_k\beta_{s_{t+1}} & \text{for } k=K+1\end{cases} \qquad (7)$$

  - **Resample $\boldsymbol{\beta}$:**

$$(\beta_1,\dots,\beta_{K+1})\sim\text{Dirichlet}(m_1,\dots,m_K,\gamma) \qquad (8)$$

- **Definitions:**
  - $\boldsymbol{\Theta}=(\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_K)$ are emission parameters,
  - $\delta_{\boldsymbol{\theta}_j}$ denotes an atom at $\boldsymbol{\theta}_j$
  - $\alpha>0$ is the local concentration parameter
  - $\gamma>0$ is the global concentration parameter
  - H is the global base distribution over the component parameters of HMM
  - f is the observation model
  - $\pi_{ij}=P(s_t=j|s_{t-1}=i)$, the elements of the transition matrix $\boldsymbol{\pi}$
  - $\boldsymbol{x}_{1:T}=(\boldsymbol{x}_1,\dots,\boldsymbol{x}_t,\dots,\boldsymbol{x}_T)$ is the sequence of observations
  - $\boldsymbol{x}_{t'}$ denotes all observations except $\boldsymbol{x}_t$
  - $s_{1:T}=(s_1,\dots,s_t,\dots,s_T)$ is the sequence of states
  - $s_{t'}$ denotes all observations except s_t
  - $n_{i,j}$ is the number of transitions from state i to state j excluding the time steps $t-1$ and t
  - $n_{i,\cdot}$ stands for the total transitions from state i
  - K is the number of states in $s_{t'}$
  - $\beta_{K+1}=\sum_{k=K+1}^{\infty}\beta_k$
  - m_k denotes the number of times the transition to state k has been drawn from the global DP.

Context Mapping

- Using the multinomial naïve Bayes classifier to map the state indicators to the binary labels: 1 for adherence, and 2 for violation.
- Feature vector of the t^{th} observation, $\boldsymbol{\rho}_t=(\rho_{t,1},\dots,\rho_{t,K})$, is a histogram

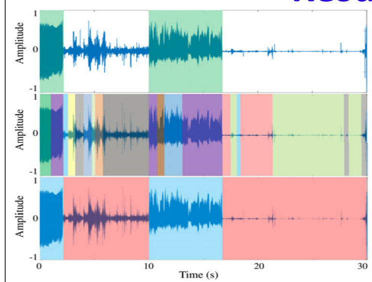
$$\hat{y}=\underset{y\in\{1,2\}}{\arg\max}\left(\log P(\tilde{y}=y)+\sum_{k=1}^{K}\rho_k\log(p_{k,y})\right) \qquad (9)$$

- $p_{k,y}$ is the probability of the k^{th} attribute being in class $\tilde{y}\in\{1,2\}$
- $P(\tilde{y})$ is the prior class probability

Experimental Setup

- **Data set:** 100 sustained vowel /a/ of length 20 s from Smartphone-PD data set
- **Acoustic features:** 12 MFCCs + frame energy
- **iHMM:** Conjugate normal-gamma prior over the Gaussian state parameters, hyper-parameters $\alpha=10$ and $\gamma=10$

Results



Comparison of the baseline systems and the proposed method. Results are in the form of mean ± STD (in %).
(TPR: true positive rate, TNR: true negative rate, Acc: Accuracy)

Method	TPR	TNR	Acc
VAD	84	96	90
NPSAR [1]	88 ± 9	91 ± 9	89 ± 8
Proposed	**97 ± 2**	**96 ± 4**	**96 ± 2**

Top: Segments hand-labeled as adhering to the protocol.
Middle: Generated states in different colors.
Bottom: Mapping state indicators to adherence label (in blue), and violation (in red).

Conclusion

- The proposed algorithm is based on splitting signals into variable duration segments by fitting an infinite HMM to the frames of signals, and identifying segments that adhere to the voice test protocols by applying a classifier.
- Using a small amount of hand-labeled data, the proposed approach can accurately identify protocol violations with a 0.2 second resolution.

References

[1] R. Badawy, Y. P. Raykov, L. J. W. Evers, B. R. Bloem, M. J. Faber, A. Zhan, K. Claes, and M. A. Little, "Automated quality control for sensor based symptom measurement performed outside the lab," *Sensors*, vol. 18, no. 4, 2018.