AUTONOMOUS CONTROL & INFO TECH

Sensitivity Analysis of Hidden Markov Models

Seifemichael Amsalu
(PhD Candidate)

Advisor:-
Abdollah Homaifar, PhD

North Carolina Agricultural and Technical State University

Outline

- ❑ Introduction
- ❑ Literature Review
  - ❑ Hidden Markov Models (HMMs)
  - ❑ Sensitivity Analysis in HMMs
- ❑ Methods
- ❑ Proposed Method
- ❑ Conclusion
- ❑ Future Works

Introduction

- Sensitivity analysis is a general technique for investigating the robustness of the output of a mathematical model
- Our focus is on parameter sensitivity analysis, which is a standard technique for studying how the output of a model varies with variation of its parameters
- Used to identify those parameters for which an accurate assessment seems important
- The results used as a basis for parameter tuning, as well as for studying the robustness of the model output to changes in the parameters
- Here the focus is on enhancing techniques for *sensitivity analysis in HMMs*, using results from research in *Bayesian networks*

HMMs

- One of the popular methods for modeling sequential and/or temporal data
- Simple enough that you can actually estimate the parameters from data and efficiently made inference
- Rich enough to handle real world applications

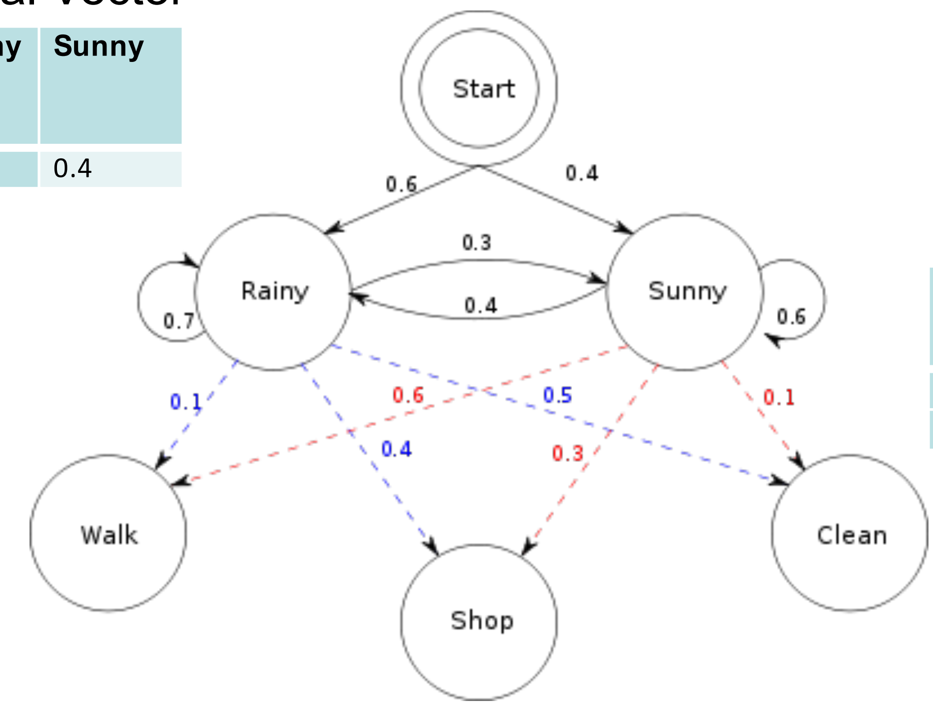
HMMs

Initial Vector

Rainy	Sunny
0.6	0.4

Transition Matrix

	Rainy	Sunny
Rainy	0.7	0.3
Sunny	0.4	0.6



Observation Matrix

	Walk	Shop	Clean
Rainy	0.1	0.4	0.5
Sunny	0.6	0.3	0.1

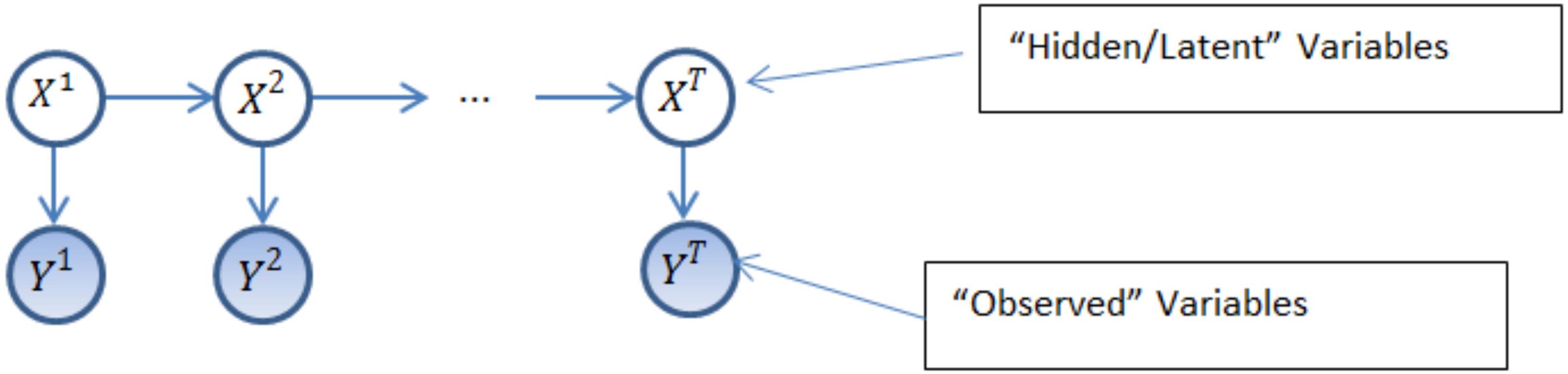
HMMs

In HMM, we have random variables as:

$$X^1, \dots, X^T \in \{1, \dots, n\}$$

$$Y^1, \dots, Y^T \in \boldsymbol{Y}\ (e.g.\ discrete, \mathbb{R}, \mathbb{R}^d)$$

It respects the graph (Trellis diagram):



HMMs

Factorization of the graphical model: joint distribution

$$p(X^1, \dots, X^T, Y^1, \dots, Y^T) = p(X^1)p(Y^1|X^1)\prod_{t=2}^{T} p(X^t|X^{t-1})p(Y^t|X^t)$$

Parameters:

Transition Matrix

- Transition Probabilities: $A(i,j) = p(X^{t+1} = j | X^t = i) \;\; (i,j \in \{1, \dots, n\})$
- Emission Probabilities: $O_i(y) = p(y|X^t = i) \; for \; i \in \{1, \dots, n\}, y \in \boldsymbol{Y})$ (i.e. O_i is a probability distribution on $\boldsymbol{Y}$)

  Density (pdf)

  $O_i(y) = p(Y^t = y|X^t = i) - - - - pmf$
- Initial Distribution: $\gamma(i) = p(X^1 = i) \;\; (i \in \{1, \dots, n\})$

$$p(X^1, \dots, X^T, Y^1, \dots, Y^T) = \gamma(i) O_{X^1}(Y^1) \prod_{k=2}^{n} A(X^{t-1}, X^t) O_{X^t}(Y^t)$$

HMMs

- **Forward-Backward (F/B) Algorithm**

Assume: - $p(Y^t|X^t), p(X^t|X^{t-1}), p(X^1)$ known

F/B: Compute $p(X^t|y)$ $y = (Y^1, \ldots, Y^T)$

Notation: - $y^{i:j} = (Y^i, Y^{i+1}, \ldots, Y^j)$, $y = Y^{1:T}$, $y^{t+1:T} = (Y^{k+1}, \ldots, Y^T)$

Forward Algorithm: Compute $p(X^t, y^{1:t})$ $\forall\, t = 1, \ldots, T$

Backward Algorithm: Compute $p(y^{t+1:T}|X^t)$ $\forall\, t = 1, \ldots, T-1$

$$p(X^t|y) \underbrace{\propto}_{X^t} p(X^t, y) = p(y^{t+1:T}|X^t, y^{1:t})p(X^t, y^{1:t}) = \overbrace{p(y^{t+1:T}|X^t)}^{Backward\ A.}\overbrace{p(X^t, y^{1:t})}^{Forward\ A.}$$

HMMs

- **Forward-Backward (F/B) Algorithm (Contd.)**

What you can do?

- Inference: $p(X^t \neq X^t|y)$ "change detection"
- Estimate parameters coupled with "Baum-Welch" Algorithm
- Sampling from posterior distribution $X|y$ --- Viterbi Algorithm

HMMs

- **Forward Algorithm**

Goal: Compute $p(X^t, y^{1:t})$

$$\alpha_t(X^t) = p(X^t, y^{1:t}) = \sum_{X^{t-1}=1}^{n} p(X^t, X^{t-1}, y^{1:t})$$

$$= \sum_{X^{t-1}=1}^{n} p(Y^t | X^t, X^{t-1}, y^{1:t-1}) p(X^t | X^{t-1}, y^{1:t-1}) p(X^{t-1}, y^{1:t-1})$$

$$= \sum_{X^{t-1}=1}^{n} \underbrace{p(Y^t | X^t)}_{Transition\ Prob.} \underbrace{p(X^t | X^{t-1})}_{Emissin\ Prob.} \underbrace{p(X^{t-1}, y^{1:t-1})}_{\alpha_{t-1}(X^{t-1})}$$

HMMs

- **Forward Algorithm (Contd.)**

$$\alpha_t(X^t) = \sum_{X^{t-1}=1}^{n} p(Y^t|X^t)p(X^t|X^{t-1})\alpha_{t-1}(X^{t-1})\, for\ t = 2, \ldots, T$$

$$\alpha_1(X_1) = p(X^1, Y^1) = p(X^1)p(Y^1|X^1)$$

$$\alpha_1, \alpha_2, \alpha_3 \ldots\ \alpha_n$$

$$\Theta(n)\, foreach\ X^t, each\ t$$

$$\Theta(T^2) foreach\ t$$

$$\Theta(nT^2) computational\ complexity\ of the\ algorithm$$

HMMs

- **Backward Algorithm**

Goal: Compute $p(y^{t+1:T}|X^t)\ \forall\, t = 1, \ldots, T-1 \ and\ \ \forall\, X^t = 1, \ldots, n$

$$\beta_t(X^t) = p(y^{t+1:T}|X^t) \ = \sum_{X^{t+1}=1}^{n} p(y^{t+1:T}, X^{t+1}|X^t)$$

$$= \sum_{X^{t+1}=1}^{n} p(y^{t+2:T}|X^{t+1}, X^t, y^{t+1})\ p(y^{t+1}|X^{t+1}, X^t) p(X^{t+1}|X^t)$$

$$= \sum_{X^{t+1}=1}^{n} \underbrace{p(y^{t+2:T}|X^{t+1})}_{\beta_{t+1}(X^{t+1})} \underbrace{p(y^{t+1}|X^{t+1})}_{Transition\ Prob.} \underbrace{p(X^{t+1}|X^t)}_{Emission\ Prob.}$$

$$\beta_t(X_t) = \sum_{X_{k+1}=1}^{n} \beta_{t+1}(X^{t+1})\ p(y^{t+1}|X^{t+1}) p(X^{t+1}|X^t)\ for\ t = 1, \ldots, T-1$$

$$\beta_T(X^T) = 1\ \forall\, X^T$$

$$\Theta(nT^2) computational\ complexity\ of the\ algorithm$$

Sensitivity Analysis in HMMs

- Usually performed by means of a perturbation analysis where a small change is applied to the parameters, upon which the output of interest is re-computed [3], [4]
- Recently, it was demonstrated that the relation between model parameters and output probabilities in HMMs can also be described by simple mathematical functions, similar to Bayesian network sensitivity functions [5]
- For determining these functions for HMMs, however, no algorithms exist

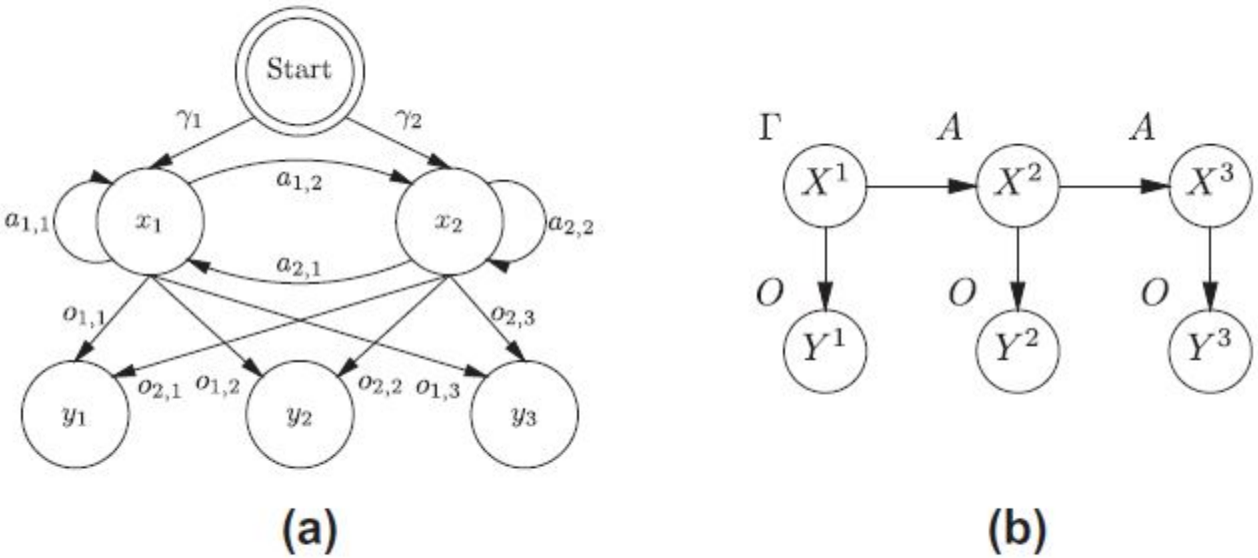
- [3] P.-A. Coquelin, R. Deguest, R. Munos, Sensitivity analysis in HMMs with application to likelihood maximization, in: Advances in Neural Information Processing Systems, vol. 22, 2009, pp. 387395.
- [4] A.Yu. Mitrophanov, A. Lomsadze, M. Borodovsky, Sensitivity of hidden Markov models, Journal of Applied Probability 42 (2005) 632642.
- [5] Th. Charitos, L.C. van der Gaag, Sensitivity properties of Markovian models, in: Proceedings of Advances in Intelligent Systems Theory and Applications Conference (AISTA), Luxembourg, IEEE Computer Society, 2004.

Sensitivity Analysis in HMMs

- It was suggested to represent the HMM as a dynamic Bayesian network, unrolled for a fixed number of time slices, and to apply existing Bayesian network sensitivity analysis algorithms [5]
- In [6] argue that these methods are inefficient for the purpose of computing HMM sensitivity functions, due to the fact that the repetitive character of the HMM, with the same parameters occurring for each time step, is not exploited in the computation
- In [6] two new algorithms are introduced, that build on existing algorithms for HMMs in order to compute the constants of the HMM sensitivity function
- In addition, It presents a new algorithm that is specially tailored to the computation of sensitivity functions directly from HMMs

- [6] S. Renooij, Efficient sensitivity analysis in hiddenMarkovmodels, in: PetriMyllymaki, Teemu Roos, TommiJaakkola, (Eds.), Proceedings of the Fifth European Workshop on Probabilistic Graphical Models, HIIT Publications 2010-2, Helsinki, 2010, pp. 241248.

Sensitivity Analysis in HMMs



- Fig. 1. A hidden Markov model representation (a) and its dynamic Bayesian network representation, unrolled for three time slices (b) [6]

- [6] S. Renooij, Efficient sensitivity analysis in hiddenMarkovmodels, in: PetriMyllymaki, Teemu Roos, TommiJaakkola, (Eds.), Proceedings of the Fifth European Workshop on Probabilistic Graphical Models, HIIT Publications 2010-2, Helsinki, 2010, pp. 241248.

Methods

The Coefficient-Matrix-Fill procedure

- Designed to compute the coefficients of the polynomial sensitivity function in Eq. (1) for transition and observation parameters
- Constructs a set of matrices containing these coefficients for each hidden state and each time slice is designed

$$p(X_v^t, \boldsymbol{y}_{\boldsymbol{e}}^{1:T})(\theta) = \sum_{i=0}^{N} c_i . \theta^i \qquad \textbf{(1)}$$

Methods

The Coefficient-Matrix-Fill procedure : Basic Idea

- The sensitivity functions for a filtering probability $p(X_v^t, \boldsymbol{y}_{\boldsymbol{e}}^{1:t})(\theta)$, as shown in Eq. (1) needs to establish coefficients $c_{v,j}^t, j = 0, \ldots, N$, where $N = t - 1$ for a transition parameter θ_a and $N = T = t$ for an observation parameter θ_o
- To compute these coefficients, a series of "Forward" matrices $F^k, k = 1, \ldots, N + 1$, are constructed with the following properties:
  - » Each matrix F^k has size $n \times k$ for $\theta = \theta_a$, or size $n \times (k + 1)$ for $\theta = \theta_o$
  - » A row i in F^k contains all coefficients for the function $p\left(X_i^t, \boldsymbol{y}_{\boldsymbol{e}}^{1:k}\right)(\theta)$
  - » A column j in F^k contains all coefficients of the $(j - 1)^{\text{th}}$-order terms of the n polynomials (one for each hidden state)

Methods

The Coefficient-Matrix-Fill procedure : Basic Idea (Contd.)

- In fact computes the coefficients for the sensitivity functions for $all\ n$ hidden states and all time slices up to and including t
- The sensitivity functions for a smoothing probability, requires the computation of a series of “Backward” matrices B^k, in addition to the forward matrices for the filter component as shown in Eq. (2)
  - » $\mathbf{p}(\mathbf{X}_{\mathbf{v}}^{\mathbf{t}}, \mathbf{y}_{\mathbf{e}}^{\mathbf{1:T}}) = \mathbf{p}(\mathbf{X}_{\mathbf{v}}^{\mathbf{t}}, \mathbf{y}_{\mathbf{e}}^{\mathbf{1:t}}, \mathbf{y}_{\mathbf{e}}^{\mathbf{t+1:T}}) = \mathbf{p}(\mathbf{y}_{\mathbf{e}}^{\mathbf{t+1:T}} | \mathbf{X}_{\mathbf{v}}^{\mathbf{t}}). \mathbf{p}(\mathbf{X}_{\mathbf{v}}^{\mathbf{t}}, \mathbf{y}_{\mathbf{e}}^{\mathbf{1:t}})$ **(2)**
- Matrices B^k will serve to compute the coefficients of the function $p(\boldsymbol{y}_{\boldsymbol{e}}^{t+1:T} | x_v^t)(\theta)$

Methods

The Coefficient-Matrix-Fill procedure (Contd.)

COEFFICIENT-MATRIX-FILL $(x_v^t, \boldsymbol{y}_e^{1:T}, \theta, H)$:

if $\theta = \theta_a$ **then** $N \leftarrow t - 1$ **else** $N \leftarrow T$

for $k = 1$ to $N + 1$ **do** $F^k \leftarrow INIT - FORWARD(\theta, K, H)$

if $T > t$ **then for** $k = t$ **to** T

do $B^k \leftarrow INIT - BACKWARD(K, H)$

for $k = 2$ **to** $N + 1$ **do** $F^k \leftarrow FILL - MATRIX(\theta, F^{k-1}, A, O)$

if $T > t$ **then for** $k = T - 1$ **downto** t

do $B^k \leftarrow FILL - MATRIX(\theta, B^{k+1}, A, O)$

Return $F^1, \ldots, F^{N+1}, B^t, \ldots, B^T$

- **Fig. 1.** High-level summary of the Coefficient-Matrix-Fill procedure

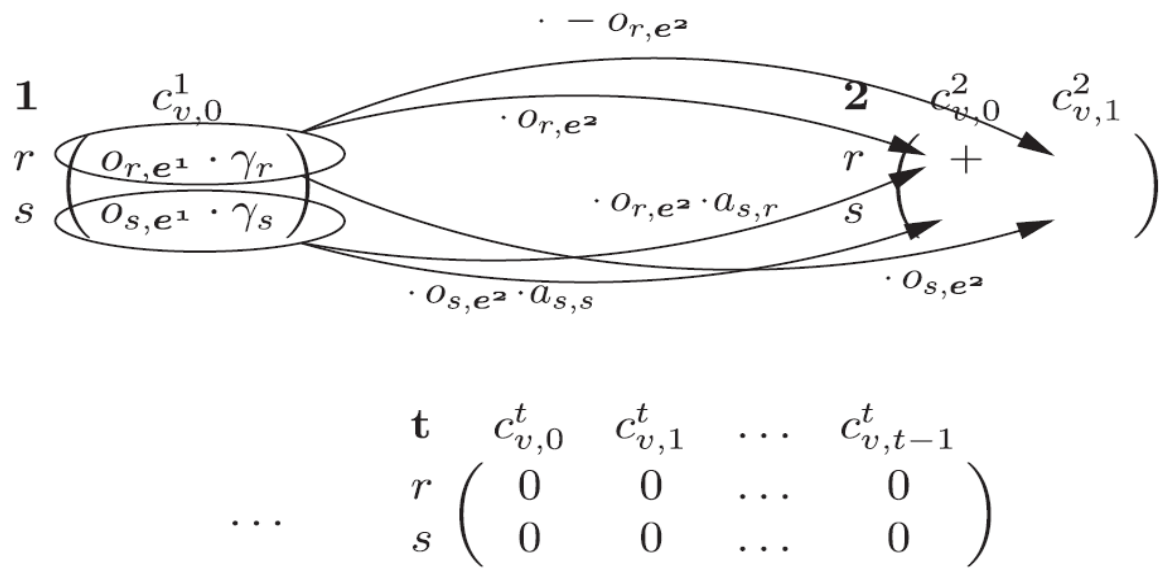
Methods

The Coefficient-Matrix-Fill procedure : Initialization and fill operations

- As summarized in Fig. 1, the procedure starts by filling the entries of matrix F^1 in accordance with the $\mathrm{t}=1$ case in the recursive expression for filter probabilities (Eq. (3)); matrix B^T is filled with all 1's
  - » $F(z,t) = p(X_v^t, y_e^{1:t})$
  - » $$F(z,t) = \begin{cases} o_{v,e^1}.\gamma_v & if\ t=1 \\ o_{v,e^t}.\sum_{z=1}^{n} a_{z,v}.F(z,t-1) & if\ t>1 \end{cases} \quad \textbf{(3)}$$
- All other matrices F^k, k>1, and $\mathrm{B}^\mathrm{k}, \mathrm{t} \le \mathrm{k} \le \mathrm{T}$, are initialized with zeroes and subsequently filled with their correct contents by the procedure

Methods

The Coefficient-Matrix-Fill procedure (Contd.)



- **Fig. 2** An example of transitioning from matrix F^1 to F^2 in the coefficient-matrix-fill procedure; here constants of the sensitivity function relating a filter probability to a transition parameter $\theta_a = a_{r,s}$, are computed [6]

Methods

The Coefficient-Matrix-Fill procedure : Sensitivity of filtering to transition parameter variation

- Here consider, the sensitivity functions $p(x_v^t, \boldsymbol{y}_{\boldsymbol{e}}^{1:t})(\theta_a)$ for a filter probability and transition parameter $\theta_a = a_{r,s}$ with n=2 hidden states and m=2 observations
- From the recursive expression for filter probabilities (Eq. (3)), it follows that for $t = 1$ we have the constant $p(x_v^t, \boldsymbol{y}_{\boldsymbol{e}}^{1:t})(\theta_a) = \mathrm{O}_{v,\boldsymbol{e}^1}.\gamma_v$, and for $t > 1$,

$$p(x_v^t, \boldsymbol{y}_{\boldsymbol{e}}^{1:t})(\theta_a) = \mathrm{O}_{v,\boldsymbol{e}^t}.\sum_{z=1}^{2} \mathrm{a}_{z,v}(\theta_a).\,p(x_z^{t-1}, \boldsymbol{y}_{\boldsymbol{e}}^{1:t-1})(\theta_a)$$

Methods

The Coefficient-Matrix-Fill procedure : Sensitivity of filtering to transition parameter variation (Contd.)

- Recall that $\theta_a = a_{r,s}$; therefore, in the above formula, $a_{r,v}(\theta_a)$ equals θ_a for $v = s$ and $1 - \theta_a$ for $v \neq s$; $a_{z,v}$ for $z \neq r$ is independent of θ_a. As a result we conclude that for $t > 1$, $p(x_v^t, \boldsymbol{y}_{\boldsymbol{e}}^{1:t})(\theta_a) =$

$$\begin{cases} O_{v,\boldsymbol{e}^t}.\theta_a.p(x_r^{t-1}, \boldsymbol{y}_{\boldsymbol{e}}^{1:t-1})(\theta_a) + O_{v,\boldsymbol{e}^t}.a_{\bar{r},v}.p\left(x_{\bar{r}}^{t-1}, \boldsymbol{y}_{\boldsymbol{e}}^{1:t-1}\right)(\theta_a) & if\ v = s \\ O_{v,\boldsymbol{e}^t}.(1-\theta_a).p(x_r^{t-1}, \boldsymbol{y}_{\boldsymbol{e}}^{1:t-1})(\theta_a) + O_{v,\boldsymbol{e}^t}.a_{\bar{r},v}.p\left(x_{\bar{r}}^{t-1}, \boldsymbol{y}_{\boldsymbol{e}}^{1:t-1}\right)(\theta_a) & if\ v \neq s \end{cases}$$

Where $\bar{r}$ denotes the state of X other than r

Methods

The Coefficient-Matrix-Fill procedure : Sensitivity of filtering to transition parameter variation (Contd.)

- ***Fill contents: initialization***. The $nx1$ matrix F^1 is initialized by setting,

  $$f_{i,1}^{1} = 0_{i,e^1} . \gamma_i$$

  for $i = 1,2$

- The remaining matrices F^k of size $nxk, 2 \leq k \leq t,$ are initialized by filling them with *zeros*
- ***Fill contents:*** $\boldsymbol{F^1, k = 2, \ldots, t}$. Column j of matrix F^k should be filled using elements from the j^{th} column of F^{k-1} that are summed or multiplied with a constant, and elements from the $(j-1)^{th}$ column of F^{k-1} that are multiplied with θ_a

Methods

The Coefficient-Matrix-Fill procedure : Sensitivity of filtering to transition parameter variation (Contd.)

- More specifically position j in row i of matrix F^k, $f_{i,j}^k$ is filled with

- $$f_{i,j}^k = \begin{cases} O_{i,e^k}.\, a_{\bar{r},i}.\, f_{\bar{r},j}^{k-1} & if\ i = s\ and\ j = 1 \\ O_{i,e^k}.\, (f_{r,j-1}^{k-1} + a_{\bar{r},i}.\, f_{\bar{r},j}^{k-1}) & 1 < j < k \\ O_{i,e^k}.\, f_{r,j-1}^{k-1} & j = k \\ O_{i,e^k}.\, (f_{r,j}^{k-1} + a_{\bar{r},i}.\, f_{\bar{r},j}^{k-1}) & if\ i \neq s\ and\ j = 1 \\ O_{i,e^k}.\, (-f_{r,j-1}^{k-1} + f_{r,j}^{k-1} + a_{\bar{r},i}.\, f_{\bar{r},j}^{k-1}) & 1 < j < k \\ -O_{i,e^k}.\, f_{r,j-1}^{k-1} & j = k \end{cases}$$

Methods

The Coefficient-Matrix-Fill procedure : (Contd.)

- **Example:** Consider an HMM with binary-valued hidden state X and binary-valued evidence variable Y. Let $\Gamma = [0.20, 0.80]$ be the initial vector for X^1, and let transition matrix A and observation matrix O be as follows:

$$A = \begin{bmatrix} 0.95 & 0.05 \\ 0.15 & 0.85 \end{bmatrix} \text{ and } O = \begin{bmatrix} 0.75 & 0.25 \\ 0.90 & 0.10 \end{bmatrix}$$

- Suppose we are interested in the sensitivity functions for the two states of X^3 as a function of transition parameter $\theta_a = a_{2,1} = p(x_1^t | x_2^{t-1}) = 0.15$, for all $t > 1$
- Suppose the following sequence of observations is obtained: $y_{\mathbf{2}}^1$, $y_{\mathbf{1}}^2$ and $y_{\mathbf{1}}^3$

Methods

The Coefficient-Matrix-Fill procedure : (Contd.)

- To compute the coefficients for the sensitivity functions, the following matrices are constructed by the Coefficient-Matrix-Fill procedure:

$$F^1 = \begin{bmatrix} o_{1,2}.\gamma_1 \\ o_{2,2}.\gamma_2 \end{bmatrix} = \begin{bmatrix} 0.25.0.20 \\ 0.10.0.80 \end{bmatrix} = \begin{bmatrix} 0.05 \\ 0.08 \end{bmatrix}$$

$$F^2 = \begin{bmatrix} o_{1,1}.a_{1,1}.f_{1,1}^1 & o_{1,1}.f_{2,1}^1 \\ o_{2,1}.\left(f_{2,1}^1 + a_{1,2}.f_{1,1}^1\right) & -o_{2,1}.f_{2,1}^1 \end{bmatrix}$$

$$= \begin{bmatrix} 0.75.0.95.0.05 & 0.75.0.08 \\ 0.90.(0.08 + 0.05.0.05) & -0.90.0.08 \end{bmatrix}$$

$$= \begin{bmatrix} 0.03563 & 0.060 \\ 0.07425 & -0.072 \end{bmatrix}$$

Methods

The Coefficient-Matrix-Fill procedure : (Contd.)

$$F^3 = \begin{bmatrix} o_{1,1}.a_{1,1}.f_{1,1}^1 & o_{1,1}.(f_{2,1}^2 + a_{1,1}.f_{1,2}^1) & o_{1,1}.(f_{2,2}^2) \\ o_{2,1}.\left(f_{2,1}^1 + a_{1,2}.f_{1,1}^2\right) & o_{2,1}.\left(-f_{2,1}^2 + f_{2,2}^2 + a_{1,2}.f_{1,2}^2\right) & -o_{2,1}.(f_{2,2}^2) \end{bmatrix}$$

$$= \begin{bmatrix} 0.02538 & 0.09844 & -0.0540 \\ 0.06843 & -0.12893 & 0.0648 \end{bmatrix}$$

- We now find for example from F^3 that

$$p(x_1^3, \boldsymbol{y}_e^{1:3})(\theta_a) = 0.02538 + 0.09844.\theta_a - 0.054.\theta_a{}^2$$

and from F^2 that

$$p(x_2^2, \boldsymbol{y}_e^{1:2})(\theta_a) = 0.07425 - 0.072.\theta_a$$

Methods

The Coefficient-Matrix-Fill procedure : (Contd.)

- Likewise, by summing column entries, we can establish the coefficients for the probability of evidence functions:

$$p(\boldsymbol{y}_e^{1:3})(\theta_a) = \left(f_{1,1}^{3} + f_{2,1}^{3}\right) + \left(f_{1,2}^{3} + f_{2,2}^{3}\right).\theta_a + \left(f_{1,3}^{3} + f_{2,3}^{3}\right).\theta_a{}^2$$

and

$$p(\boldsymbol{y}_e^{1:2})(\theta_a) = \left(f_{1,1}^{2} + f_{2,1}^{2}\right) + \left(f_{1,2}^{2} + f_{2,2}^{2}\right).\theta_a$$

Methods

The Coefficient-Matrix-Fill procedure : (Contd.)

- Together these give the following sensitivity functions for two filtering tasks:

$$p(x_1^3 | \boldsymbol{y}_e^{1:3})(\theta_a) = \frac{-0.054.\theta_a{}^2 + 0.09844.\theta_a + 0.02538}{0.0108.\theta_a{}^2 - 0.03049.\theta_a + 0.09381}$$

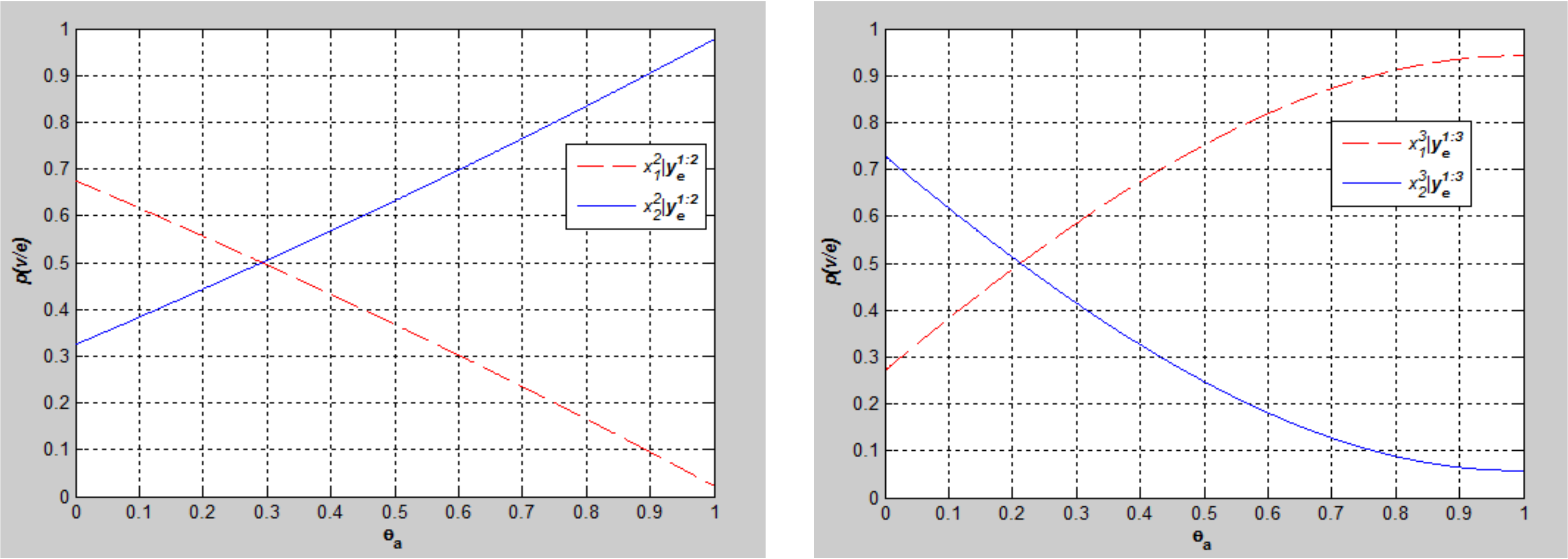
and

$$p(x_2^2 | \boldsymbol{y}_e^{1:2})(\theta_a) = \frac{-0.072.\theta_a + 0.07425}{-0.012.\theta_a + 0.10988}$$

which are displayed in Fig. 2

Methods

The Coefficient-Matrix-Fill procedure : (Contd.)



(a) (b)

- **Fig. 3**. Sensitivity functions $p(x_v^2|\mathbf{y_e}^{1:2})(\theta_a)$ for both states of X^2(a), and $p(x_v^3|\mathbf{y_e}^{1:3})(\theta_a)$ for both states of X^3(b)

Proposed Method

- In this method, to calculate the coefficients the Transition Matrix A is divided into $\bar{A}$ and $\hat{A}$ that are independent and dependent on θ_a respectively. For previous example:

$$\bar{A} = \begin{bmatrix} 0.95 & 0.05 \\ 0 & 1 \end{bmatrix} \text{ and } \hat{A} = \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix}$$

- The Observation Matrix O is represented as diagonal matrices at the time t of observation sequence
- For the observation sequence, $y_{\mathbf{2}}^1$, $y_{\mathbf{1}}^2$ and $y_{\mathbf{1}}^3$, the diagonal matrices become

$$O^1 = \begin{bmatrix} 0.25 & 0 \\ 0 & 0.10 \end{bmatrix}, O^2 = \begin{bmatrix} 0.75 & 0 \\ 0 & 0.90 \end{bmatrix}, \text{ and } O^3 = \begin{bmatrix} 0.75 & 0 \\ 0 & 0.90 \end{bmatrix}$$

Proposed Method

- Once the Transition and Observation matrices are represented as above, the sensitivity coefficients are computed with matrix multiplication as shown in the following function

F=HMMSensCoeffFilteringToTransParam$(A, O, \Gamma, \mathbf{e}, \theta)$

$\bar{A} = A$

$\bar{A}_{r,s} = o$

$\bar{A}_{r,:} = \bar{A}_{r,:}./\sum_{i=1}^{n} \bar{a}_{r,i}$

$\widehat{A} = A - \bar{A}$

$\widehat{A}_{r,:} = \widehat{A}_{r,:}./\widehat{A}_{r,s}$

$O^1 = diag(O_{:,\mathbf{e}^1})$

$F^1 = O^1 * \Gamma'$

for $k = 2$ **to** t

$O^k = diag(O_{:,\mathbf{e}^k})$

$Ftmp_1 = [O^k * \bar{A}' * F^{k-1}, zeros(n,1)]$

$Ftmp_2 = [zeros(n,1), O^k * \widehat{A}' * F^{k-1}]$

$F^t = Ftmp_1 + Ftmp_2$

end

Return $F^1, ..., F^t$

- where $\boldsymbol{e}$ is the sequence of observation and $\theta = a_{r,s}$

Proposed Method

```
e=randi([1,2],1,1000);

t1 =

    2.2741    2.1671    2.1773    2.1781    2.1656    2.1851    2.1693    2.1664    2.1710    2.1727

mean(t1)=

          2.1827

t2 =

    0.0672    0.0469    0.0455    0.0471    0.0455    0.0447    0.0469    0.0466    0.0452   0.0452

mean(t2)

          =  0.0481
```

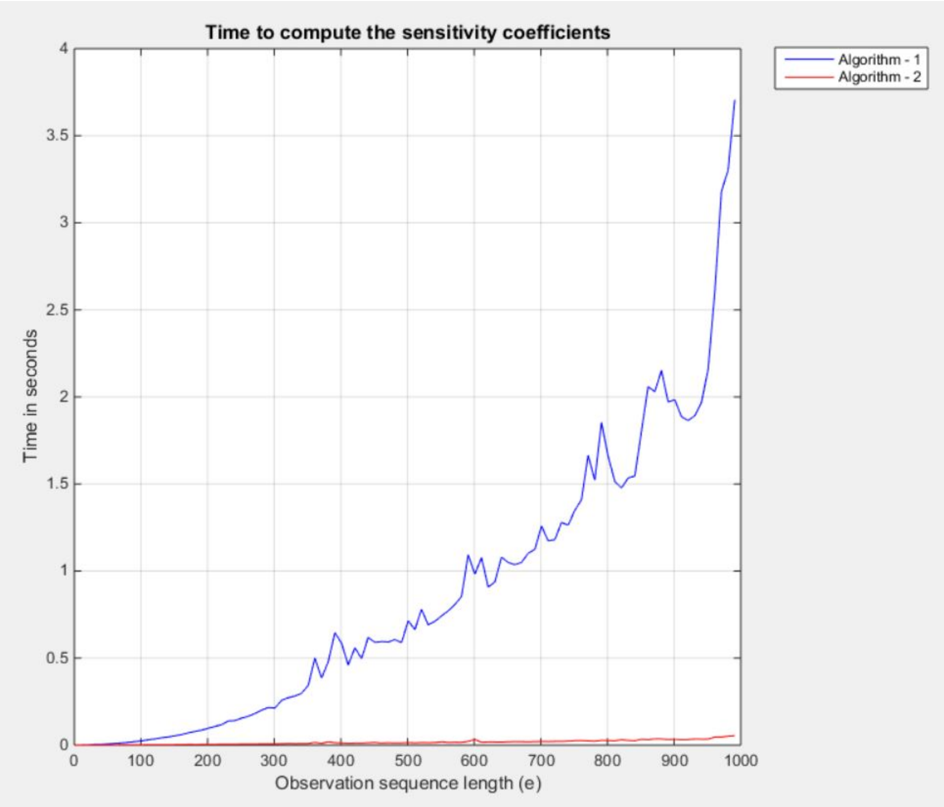
Percentage Improvement = (abs (mean (t2)-mean (t1))/mean (t1))*100 %

= 97.7975 %

Proposed Method

```
e=randi([1,2],1,10000);

t1 =

    202.5056  200.8497  200.0616  199.9931  200.5587  201.3353  200.6778  203.2773  200.3261  201.3920

mean(t1)

            =  201.0977

t2 =

    14.5198   18.7722   22.3286   19.6737   19.5887   22.6389   19.9924   17.7525   11.9295   8.6692

mean(t2)

            =   17.5866

Percentage Improvement = (abs (mean (t2)-mean (t1))/mean (t1))*100 %

                                     =   91.2547
```

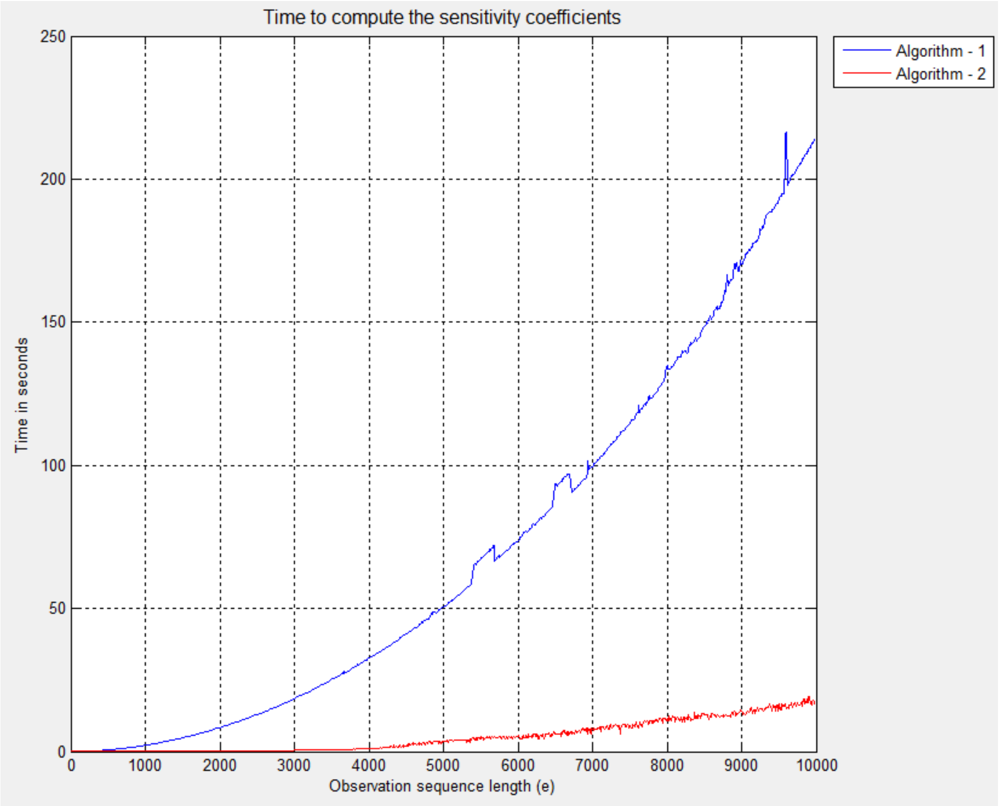
Proposed Method



- *Algorithm -1* is Coefficient-Matrix-Fill procedure
- *Algorithm-2* is Proposed Method

- **Fig. 4.** Time in seconds to compute the sensitivity coefficients for an observation sequence length from 1 to 1000 with a step size of 10

Proposed Method



- **Fig. 5.** Time in seconds to compute the sensitivity coefficients for an observation sequence length from 1 to 10000 with a step size of 10

Conclusion

- It is shown that it is more efficient to compute the coefficients for the HMM sensitivity function directly from HMMs
- The proposed method exploits the simplified matrix formulation for HMMs
- A simple algorithm that computes the coefficients for the sensitivity function for ***all hidden states*** and ***all time steps*** is presented
- It is differ from other approaches in:
  - Do not depend on a specific computational architecture
  - Do not require a Bayesian network representation of HMM
- The proposed method has shown significant improvement over coefficient matrix procedure in computational time

Future Works

- Sensitivity of filtering to observation parameter variation
- Sensitivity of smoothing to transition parameter variation
- Sensitivity of smoothing to observation parameter variation
- Sensitivity of predicted future observations $p\left(\mathbf{y}_{\boldsymbol{e}}^{t} \middle| \mathbf{y}_{\boldsymbol{e}}^{1:\mathrm{T}}\right)(\theta), T < t$
- Sensitivity of the most probable explanation (MPE) to parameter variation
- Extend the current research to sensitivity analysis in which different types of model parameter are varied simultaneously

Reference

1. H. Chan, A. Darwiche, When do numbers really matter?, Journal of Artificial Intelligence Research 17 (2002)
2. L.C. van der Gaag, S. Renooij, V.M.H. Coup, *Sensitivity analysis of probabilistic networks*, in: Advances in Probabilistic Graphical Models, Springer Series: Studies in Fuzziness and Soft Computing, vol. 213, 2007, pp. 103124
3. P.-A. Coquelin, R. Deguest, R. Munos, Sensitivity analysis in HMMs with application to likelihood maximization, in: Advances in Neural Information Processing Systems, vol. 22, 2009, pp. 387395.
4. A.Yu. Mitrophanov, A. Lomsadze, M. Borodovsky, Sensitivity of hidden Markov models, Journal of Applied Probability 42 (2005) 632642
5. Th. Charitos, L.C. van der Gaag, Sensitivity properties of Markovian models, in: Proceedings of Advances in Intelligent Systems Theory and Applications Conference (AISTA), Luxembourg, IEEE Computer Society, 2004
6. S. Renooij, Efficient sensitivity analysis in hiddenMarkovmodels, in: PetriMyllymaki, Teemu Roos, TommiJaakkola, (Eds.), Proceedings of the Fifth European Workshop on Probabilistic Graphical Models, HIIT Publications 2010-2, Helsinki, 2010, pp. 241248
7. S. Russell, P. Norvig, Artificial Intelligence: A Modern Approach, third ed., Prentice Hall, 2010
8. H. Chan, A. Darwiche, Sensitivity analysis in Bayesian networks: from single to multiple parameters, in: M. Chickering, J. Halpern (Eds.), Proceedings of the Twentieth Conference on Uncertainty in Artificial Intelligence, AUAI Press, Arlington, VA, 2004, pp. 67–75

Reference

9. Th. Charitos, Reasoning with Dynamic Networks in Practice, Ph.D. Thesis, Utrecht University, The Netherlands, 2007
10. P. Dymarski (Ed.), Hidden Markov Models, Theory and Applications, InTech Open Access Publishers, 2011
11. P.-A. Coquelin, R. Deguest, R. Munos, Sensitivity analysis in HMMs with application to likelihood maximization, in: Advances in Neural Information Processing Systems, vol. 22, 2009, pp. 387–395
12. A.Yu. Mitrophanov, A. Lomsadze, M. Borodovsky, Sensitivity of hidden Markov models, Journal of Applied Probability 42 (2005) 632–642
13. L. Rabiner and B. Juang, ”An introduction to hidden Markov models,” IEEE ASSP Mag., vol. 3, no. 1, pp. 4-16, Jan. 1986.
14. L. Rabiner, ”A tutorial on hidden Markov models and selected applications in speech recognition,” Proc. IEEE, vol. 77, no. 2, pp. 257-286, Feb. 1989.
15. L. Baum and T. Petrie, ”Statistical inference for probabilistic functions of finite state Markov chains,” Ann. Math. Stat., vol. 37, no. 6, pp. 1554-1563, Dec. 1966.

?