[研究新領域報導]

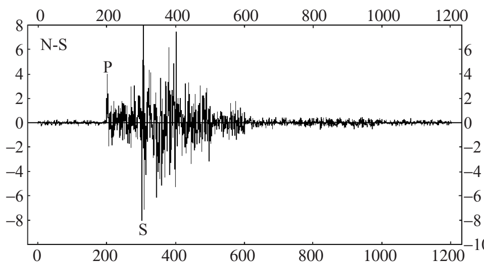
隱藏式馬可夫模型之轉折點偵測

中央大學統計研究所暨
中央研究院統計科學研究所 傅承德
中央研究院統計科學研究所 羅盛豐

一、引言

隨機動態系統之轉折點變動的偵測問題在各種不同的領域中有廣泛的應用，其中包括工業品質控管、訊號分割、財務工程、生物醫學訊號處理，DNA 資料組之轉折點偵測，以及圖案之邊緣偵測。所謂突然變動意即在瞬間某一種性質突然的改變，但之前或之後該性質皆是定態的。因突然變動不會有很大幅度的變化，很多變動檢測問題都是在討論微小變動的偵測。如上所述，大部分在測度之內的資訊都是非定態，並且突然變動的檢測在許多應用上皆很有價值。故我們需要尋求一些工具來探測其是否有此變動產生。而這些問題都可以用一參數值在不特定時間內會突然變化的統計模型加以描述。

在訊號分割應用中最典型的問題即是利用訊號的自動分割做爲處理的第一步，然後利用一分割演算法將同質的訊號獨自分離爲一層，而每一層的長度視分析訊號性質的差異而不同。在標準轉折點偵測問題的結構中，會有一系列在某一未知時間會改變其結構的觀察值，而我們的目標是即時檢測出這些改變。自動分割演算法已被引進至連續語音訊號的自動分析，及生物醫學訊號之自動分析中而成爲一相當有效的工具。圖一中展示了一個對連續語音訊號自動分割的範例。



圖一 Automatic segmentation of a continuous speech signal

現實世界中所產生的現象通常都可以轉化成訊號。而這些訊號在本質上可以是連續的（如音樂、語音樣本、溫度量測等），或者是離散的（如有限符號系統、電碼書的量子化向量、DNA 形式之資料組等）。而資料的來源可以是定態的（即其統計性質不隨時間變化），或是非定態的（即統計性質會隨時間變化）。此訊號可以很單純（即它們完全來自於同一來源），或者可能受到其他訊號來源的污染（如噪音），或者在傳送時受到扭曲、反射、壓縮等。這其中的一個最根本問題即爲如何將眞實世界的訊號轉化成一訊號模型。我們用以下幾個理由說明爲何會對應用訊號模型有興趣。第一，訊號模型可以理論性的描述那種訊號處理系統能產生特定的訊號。舉例來說，如果我們有興趣的是一受噪音與傳輸干擾的語音訊號,我們可以使用訊號模型來設計一個能移除噪音與傳輸干擾的系統。第二，訊號模型可以讓我們在訊號來源未知的情況下對其有一定程度的了解。而此性質在需付出高成本以獲得訊號來源時是特別重要。在此情況下，若有一個完善的訊號模型，則我們可以模擬並儘可能的依此以了解訊號來源。最後也是最重要的理由是訊號模型在現實應用中通常皆運作良好，並讓我們很有效率的了解重要的應用系統，如預測系統、辨識系統、鑑別系統等。

接下來我們介紹一些重要類型的訊號模型，並用它來捕捉特定訊號的性質。一般來說我們可以將訊號模型分爲確定性模型與統計模型兩大類。確定性模型通常探討訊號之某些已知、特定的性質。在這種情況下，訊號模型的特性就是平舖直敘的；唯一的需要即爲估計此訊號模型的參數。而統計模型則是嘗試去捕捉訊號的統計特性。其基本假設爲訊號可以被一具參數之隨機過程所描述，並且其參數是可以被精準地估計

的。在此我們介紹一種主要隨機訊號模型，即隱藏式馬可夫模型(hidden Markov models, HMM)。一隱藏式馬可夫模型是一雙層的隨機過程，有一隱含的隨機過程是無法被直接觀察的，但我們可以藉由另一組產生一系列的觀察值來觀察其變化。此一模型受到學界歡迎的原因有二：第一，此模型有豐富且嚴謹的數學結構，因而能在應用上提供多方面的理論基礎。第二，當應用得當時，此模型在許多不同實際領域上皆運作良好，如語音辨識、分子生物學、經濟學與生物資訊。

Shewhart 的三倍標準差控制圖(3-sigma control chart)在轉折點檢測中為一開創性的工作、它是在監測一控制製造流程定態性之簡單並且重要的工具。而更有效率的轉折點檢測法為累積和程序(cumulative sum procedure, CUSUM)及 Shiryayev-Roberts-Pollak (SRP)演算法。在一變化之前與變化之後為獨立觀察值的簡單系統中，SRP 演算法是由 Shiryayev [13]所提出，其目標為在固定誤警機率上限下，極小化其延遲期望值。Roberts [12]考慮非貝式設定，並以模擬方法研究此法則的平均延遲時間，並發現其表現甚佳。Pollak [10]證明在修正後的 Shiryayev-Roberts 法則具有漸近最優性。Basseville and Nikiforov [1] 在其書中表示現有文獻中，雖對複雜系統之檢測演算法已有許多研究，但對模型上的統計性質與檢測步驟之最優性卻只有少量的研究。本研究的主要目的即為探討在隱藏式馬可夫模型中 SRP 轉折點檢測法則在統計理論方面的性質。我們證明了其在一合理範圍的轉折點偵測法則中的漸進最優性。因對隱藏式馬可夫模型統計分析的需要，我們也證明了一些機率論上的相關結果。結果詳見 Fuh [4, 5]。第二節討論隱藏式馬可夫模型之結構性改變的偵測。第三節討論在訊號有被壓縮的情況及未來的研究方向。

二、隱藏式馬可夫模型結構性改變之偵測

Mei [9]中彰顯出一族特定的隱藏式馬可夫模型其結構性改變的問題，其 HMM 轉折改變點是被定義發生在一已知的狀態(state)下。然而，我們並不是完全地清楚了解此定義的一般性，即便在狀態數 $K = 2$ 的情況下已被完全地說明。事實上，這個對於改變後狀態被考慮為是一個吸收狀態(absorbing state)的想法令人感到一點點困惑，因為考慮這些狀態屬於可溝通性狀態比他們身為一個收縮狀態的必要性還來的重要。舉例來說，考慮以下這種機率轉移矩陣的模型：

$$\begin{pmatrix} p & 1-p \\ q & 1-q \end{pmatrix} \; 0 < p, q < 1,\ p \neq q. \qquad (1)$$

此架構違反了狀態二是一個收縮狀態的想法,因此在這樣的例子中考慮結構性改變的問題就變地有趣。假設這裡給定如 Mei [9]中的限制：$\pi_1^0 = 1$即一馬可夫鏈從狀態一開始，然後到了某個給定的轉折時點後，此鏈會被保證一定屬於狀態二。然而，此鏈經過狀態二的事件將會不時地發生(infinitely often)，而且由於此例 HMM 不可化簡的特性，此模型在當此一轉折時點存在或是不存在時都會有相同的極限分佈(limiting distribution)。事實上，即便狀態二為一吸收狀態($q = 0$)，相同地也成立。因此，考慮一不可化簡的隱藏式馬可夫鏈似乎為一基本的要求，更進一步來講，此鏈的可溝通性類狀態必須在改變前和改變後的初始分佈下皆有所區別。

在給定不可化簡性和可溝通類型的考量限制下，以下也許提供了一個有關 HMM 轉折點問題更直覺的了解和看法。考慮兩個模型，一個是在 $\mathbf{P}_f$下的 HMM，另一個是在 $\mathbf{P}_g$ 獨立且同態的模型，而此兩模型以轉折時點作為區隔。在這個結構性改變的定義下，一個馬上就會產生的問題就是，在 $\mathbf{P}_g$ 也是一個 HMM 的架構之下，此定義是否依然有效。當然此一新的 HMM 有著與原先不一樣的結構，不論是在參數或轉移機率的不同，或是在狀態空間狀態數目的不同。這樣的考量下可能就會是有關不可化簡隱藏式馬可夫鏈轉折點問題的一個一般性的定義。

為了闡述這個想法，我們仔細研究了一個兩狀態的隱藏式馬可夫模型，其機率轉移矩陣型態如(1)而改變情況為 q 轉變成 0。我們所選用的參數如下：初始分配π^i = [0.5, 0.5]、p = 0.9、狀態一為 $N(0, 1)$ 分佈而狀態二為 $N(0.3, 1)$分佈。令τ表示累積和程序(CUSUM)的停止時間(stopping time)(參見 Fuh [4]式(5.1)), 並利用粒子濾波器(particle flter)來執行。對於不同的 q 值，我利用模擬來選取邊界值 b 使得錯誤警訊(false alarm)平均串聯長度(ARL, average run length)約為

800。[1] 假設轉折改變時點爲時間 1，表一是透過 1000 回模擬 2000 個觀測值來得到的：

表 1 CUSUM change point detection for structure change in HMM

q	τ(%)	b	q	τ(%)	b
0.9	10.20 (100)	0.345	0.4	63.52 (100)	2.34
0.8	13.38 (100)	0.509	0.3	91.97 (100)	3.06
0.7	20.81 (100)	0.842	0.2	139.22 (100)	3.9
0.6	30.19 (100)	1.26	0.1	237.13 (100)	4.831
0.5	45.01 (100)	1.8			

表中括弧內的數字表示眞正碰觸到邊界值的百分比率，而這個例子告訴了我們 ARL 仍舊是一個合理的評判標準。另外，在 Tartakovsky and Blanding [14]和Blanding et al. [2]的其他例子中則顯示出了在以 ARL 爲評判標準下，累積和程序(CUSUM)爲次佳的(suboptimal)。

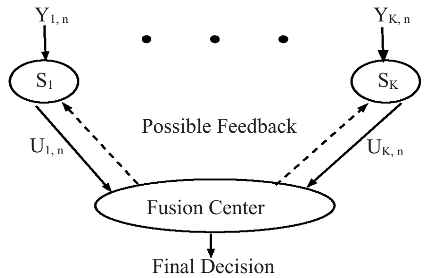
三、隱藏式馬可夫模型中分散式最速轉折點偵測的最佳穩定二元數字轉換器

分散式最速轉折點偵測的問題已在感測器網路(sensor network)中被大量地研究，此網路含有一組數個的感測器，分別從一個隱藏式馬可夫模型 **X** 收集觀測資料，然後再個別地將感測訊息傳送給一個共同的中央處理器，稱爲融合中心(fusion center)。當觀測停止時，此中心便會做出一個最後的決策。假設模型 **X** 裡頭有一參數θ，而此參數值會在某一未知的時點從θ_0轉成θ_1。而我們要處理的問題就是在誤警平均串聯長度的限制下去決定個別感測器連同融合中心的一組策略使得偵測延遲(detection delay)爲最適。Fuh and Mei [7]研究了一種連帶穩定二元感測訊息的累積和型態的融合規則並且介紹了一個簡單選取個別最佳的局部感測門檻。圖二爲一般感測網路的示意圖。

1. 問題的公式化表述

令 $\mathbf{X} = \{X_n, n \geq 0\}$是一個離散時間的馬可夫鏈定義在一個有限的狀態空間 $D = \{1,\ldots, d\}$上頭，其機率轉移矩陣爲

1 此模擬是以 Fuh [5]中的定理 6 爲依據。在獨立同態轉折點問題的架構下，則可利用 Pollak and Siegmund [11]的式子(14)來求得 b。



圖二 A general setting for sensor networks

$$P_\theta = \begin{bmatrix} p_\theta(1,1) & \cdots & p_\theta(1,d) \\ \vdots & \ddots & \vdots \\ p_\theta(d,1) & \cdots & p_\theta(d,d) \end{bmatrix}$$

又穩定分配(stationary distribution)爲

$$\pi_\theta = (\pi_\theta(1), \cdots, \pi_\theta(d))^t, \tag{2}$$

其中 t 表示矩陣的轉置。假設系統中有 K 個感測器，在時間點 n 時，每個感測器 S_k 皆會接收到一個觀測資料 $Y_{k,n}$，此資料可以被模型化爲

$$Y_{k,n} = h_k(X_n) + \varepsilon_{k,n}, \tag{3}$$

其中 $h_k(\cdot)$是一個定義在 **X** 狀態空間上的實數值函數而感測噪音(sensor noises) $\varepsilon_{k,n}$'s 爲常態 $N(0, \sigma_k^2)$隨機變數，這些隨機變數無論是對不同的時間或是對不同的感測器皆獨立。這裡的 $h_k(\cdot)$或是 σ_k^2 *可以*和也*可以不*和指標θ有關，不過爲了簡化符號我們都不再標示出來。此外，在時間點 n，爲了資料壓縮和通訊管道頻寬的限制等原因，每個S_k都會將原始感測觀測資料 $Y_{k,n}$加以量化並將此量化後的資料當成一個感測訊息 $U_{k,n}$ 傳送給融合中心。爲了簡單起見，這裡我們假設

$$U_{k,n} = \phi_{k,n}(Y_{k,n}) = \begin{cases} 1, & \text{if } Y_{k,n} \geq \lambda_k \\ 0, & \text{if } Y_{k,n} < \lambda_k \end{cases}, \tag{4}$$

其中每個限制門檻λ_k 都是由統計學家所選取的常數。接下來，融合中心就會把這些由感測器傳來的訊息流 $U_{k,n}$'s 當做輸入資料來做最後的決策。

而在有關 HMM 最速轉折點偵測的問題裡，通常假設隱藏式馬可夫模型 **X** 裡的參數值θ會在某一個未知（可能是無窮）的時點ν 從θ_0轉變成θ_1，而這樣個改變當然也會導致感測觀測資料 $Y_{k,n}$'s 分配的改變。我們的目標就是在所有可能的穩定門檻λ_k's（感測器端）和所有可能的決策

規則（融合中心端）中，去盡快地偵側到眞正的改變時點。

2. 問題討論

Fuh and Mei [7]文中提出了以下幾個有趣的議題：

1. ***任意的穩定感測器***：我們這裡只考慮了一種特別的穩定二元數字轉換器，也就是只把當下的感測觀測値拿來和一個固定的門檻來做比較。對於獨立的感測觀測資料來說，這種簡單型式的感測數字轉換器,，或者像 Tsitsiklis [15]所提的更一般的單調概似比數字轉換器(MLRQ)，都可以在所有可能的（穩定或非穩定的）感測數字轉換器中導致漸近最佳分散式偵測系統。因此在 HHM 模型之下，這樣的最適性質（至少在所有的穩定感測數字轉換器之下）是否仍然正確就變的是一個有趣的問題了。
2. ***非穩定的數字轉換器***：我們這裡只關注有著穩定二元數字轉換器的分散式偵測系統，進而在這個類別底下發展有效率的機制。一個很自然大家都可能會問的問題是，在穩定數字轉換器之下的"最佳"機制是否仍然是漸近最適的，如果我們允許感測門檻λ_k's 或者更廣泛來說，感測數字轉換器，可以對過去的感測訊息做應變（也就是說融合中心可以在每一個時間間隔改變系統內的參數）。令人意外地，Mei [8]證實了上述的猜想對於獨立的觀測資料情況下是成立的,，但是對於 HMM 模型的情況仍存有懷疑。我們將會在其他的文章中提出在 HMM 模型下相對應的漸近性質的研究。
3. ***區間長度的效果***：在實務上爲了更進一步的減少通訊成本，感測器可以被設計成每隔 M 個時點傳送一次訊息。但什麼是最好的區間長度 M 呢？這裡我們有兩個考量要取捨，一個是我們希望能快速地偵測到眞正的改變時點而另一個就是通訊成本。我們可以利用 Crow and Schwartz [3]中對於獨立觀測資料的想法來推廣至 HMM 模型下的情況。
4. ***更複雜的模型***：這裡我們只考慮了一個有限狀態空間的隱藏式馬可夫模型(3)，然而在許多的應用上，我們可能要處理一個連續的馬可夫鏈或是更複雜的隱藏式馬可夫模型 cf. Fuh [6]。直接推廣我們的結果到這些更複查的情形上頭似乎是可行的,，但是也許我們必須特別的注意 Fuh and Mei [7]中的條件。

参考文獻

[1] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*, Prentice Hall Information and System Sciences Series, Prentice-Hall (1993).

[2] W. Blanding, P. Willett, Y. Bar-Shalom, and S. Coraluppi, *ICASSP*, Honolulu, HI, April (2007).

[3] R. W. Crow and S. C. Schwartz, *IEEE Transactions on Aerospace and Electronic Systems*, **32**, 267 (1996).

[4] C. D. Fuh, *Annals of Statistics*, **31**, 942 (2003).

[5] C. D. Fuh, *Annals of Statistics*, **32**, 2305 (2004).

[6] C. D. Fuh, *Annals of Statistics*, **34**, 2026 (2006).

[7] C. D. Fuh and Y. Mei, Optimal Stationary Binary Quantizer for Decentralized Quickest Change Detection in Hidden Markov Models, *Fusion 2008 Best Paper*, to appear (2008).

[8] Y. Mei, *IEEE Transactions on Information Theory*, **51**, 2669 (2005).

[9] Y. Mei, Is Average Run Length to False Alarm Always an Informative Criterion, *Sequential Analysis*, to appear (2008).

[10] M. Pollak, Annals of Statistics, **13**, 206 (1985).

[11] M. Pollak and D. Siegmund, *Biometrika*, **72**, 267 (1985).

[12] S. W. Roberts, *Technometrics*, **8**, 411 (1966).

[13] A. N. Shiryayev, *Theory of Probability and Its Applications*, **8**, 22 (1963).

[14] A. Taktakovsky and W. Blanding, Quickest Detection of Change in Distribution for Hidden Markov Models - A Case Study, Preprint (2006).

[15] J. N. Tsitsiklis, *IEEE Transactions on Communications*, **41**, 550 (1993).