المجلة العراقية للعلوم الاحصائية (9) 2006
ص ص [67–91]

استخدام سلاسل ماركوف المخفية في تمييز حروف العلة في اللغة الإنكليزية

د.حسن محمد الياس[*] رنا بشار حسين[**]

الملخص

تمت في هذا البحث دراسة نماذج ماركوف المخفية التي هـي مجموعة منتهية من الحالات ، كل حالة تقترن بتوزيع احتمالي ، أما الانتقالات ما بين الحالات فتُحدد بوساطة مجموعة من الاحتمالات وتُسمى الاحتمالات الانتقالية ، وبشكل عام ، تتولد الحالة الناتجة (المشاهدة) طبقاً لتوزيع الاحتمالية المقترنة ، اذ توجد احتمالية ناتجة فقط ولا توجد حالة ظاهرة يمكن أن تُشاهد ولهذا فإن الحالات تكون مخفية .

تم عرض المسائل الأساسية لنماذج ماركوف المخفية (Hidden Markov Models) (HMMs) وهي :

- حساب الإمكان لمتسلسلة المشاهدات (O) عندما يكون المعطى هو النموذج $\lambda = (A , B , \pi)$ ، أي إيجاد $P(O|\lambda)$ ،حيث أن :

A = مصفوفة احتمالات انتقال الحالة .

B = مصفوفة احتمالية المشاهدات .

π = توزيع الحالة الابتدائية .

- اختيار متسلسلة الحالة المثلى لعملية ماركوف المخفية .
- إيجاد النموذج $\lambda = (A , B , \pi)$ الذي يكون فيهِ الإمكان أعظم ما يمكن ، أي إعادة تقدير معلمات النموذج لتعظيم $P(O|\lambda)$.

[*] استاذ مساعد/ قسم الاحصاء/ كلية علوم الحاسبات والرياضيات/ جامعة الموصل
[**] مدرس مساعد/ كلية الادارة والاقتصاد/جامعة الموصل

تاريخ التسلم : 29/ 8 /2005 ـــــــــــــ تاريخ القبول : 19/ 12 /2005

فضلاً عن حلولها .

Using Hidden Markov Chains in Recognition of Vowel Letters in English Language

ABSTRACT

This study deals with hidden Markov models . These models consist of sets of finite states , each one of them is associated with a probability distribution . The transition among the states is governed by a set of probabilities namely " Transition probabilities " . In general , the final observation produced according to the associated probability , where there is only probability production instead of a clear visible states . Therefore , these states are described as hidden

The basic problems for hidden Markov models (HMMs) are :

- The probability account for the observation sequence (O) when the model is given $\lambda = (A , B , \pi)$, i.e. $P(O|\lambda)$ Where :

  A = The state transition probabilities .

  B = The observation probability matrix .

  π = The initial state distribution .
- Choosing the optimal state of the sequence for hidden Markov process.
- Finding the model $\lambda = (A , B , \pi)$ which has the greatest probability , i.e. re-estimating the model to maximiz $P(O|\lambda)$.

in addition to its results .

1- مقدمة

هناك بعض الاختيارات الممكنة حول أي نوع من نماذج الإشارة هو المستخدم لتمييز خواص الإشارة ، وبشكل موسع استطاع العلماء تقسيم أنواع نماذج الإشارة إلى : نماذج التحديد والنماذج الإحصائية ، إذ إن نماذج التحديد (بشكل عام) تستخدم بعض الخواص المحددة المعروفة للإشارة ، مثلاً تمثل موجة الجيب ، ففي هذه الحالة يكون التحديد لنموذج الإشارة (بشكل عام) واضحاً ومباشراً ، وهذا مطلوب في تحديد (أو تقدير) قيم المعلمات لنموذج الإشارة (مثل : المدى ، التردد ، شكل (أو صورة) موجة الجيب) ، أما الصنف الثاني لنماذج الإشارة فهو مجموعة

من النماذج الإحصائية التي تُستخدم فقط لتمييز الخواص الإحصائية للإشارة ، ومن الأمثلة على هذه النماذج ، العمليات الكاوسية ،وعمليات بواسون وعمليات ماركوف وعمليات ماركوف المخفية ...الخ .
تُعرف عملية ماركوف المخفية على أنها عملية تصادفية مخفية مركبة وذلك لأنها عملية ضمنية لا يمكن مشاهدتها ، ولكن يمكن مشاهدتها من خلال عملية تصادفية أُخرى اذ إن الأخيرة تعطي متسلسلة من المشاهدات، فهي عملية مزدوجة ذات فضاء حالة منتهي مع عملية ماركوفية مخفية تتحكم باختيار الحالات **(Rabiner,1989)** .

دُرِست (HMMs) ، التي هي من النماذج التصادفية ، في أواخر الستينات وبداية السبعينات من القرن العشرين وقد بدت المسألة صعبة جداً في بادئ الأمر ولكن بمرور الزمن ونتيجةً للبحوث والدراسات التي جرت حول الموضوع فقد تم تسهيله نظرياً وتطبيقياً في مجالات كثيرة من مجالات الحياة .

في عام (1989) قدَّم الباحث (Rabiner) بحثاً موسعاً حول (HMMs) حاول من خلاله (وبشكل دقيق ومنظم) استعراض الجوانب النظرية لهذا النوع من الصياغة الإحصائية وإظهار كيفية تطبيقها في مسائل تمييز الكلام ، وقد عدّ هذا البحث المصدر الأساسي للعديد من البحوث الأخرى المختصة بنماذج ماركوف المخفية (Rabiner , 1989) .

في عام (1999) ألقى (Parker) محاضرة بعنوان (ملف نماذج ماركوف المخفية) تضمنت المقدمة لـ(HMMs) وكيفية بنائها (Parker , 1999) .

كذلك في عام (2001) قدَّم(Ellis) بحثاً بعنوان (تمييز المتسلسلة) أوضح فيهِ كيفية تمييز المتسلسلات الإحصائية فضلاً عن نماذج ماركوف المخفية وحالاتها وطرائق حسابها (Ellis , 2001) .

كذلك في عام (2003) قام (Zhai) بتقديم بحثٍ حول الملاحظات المختصرة والمفيدة جداً في نماذج ماركوف المخفية (Zhai , 2003) .

كذلك في عام (2003) قدَّم (Stamp) دراسة بعنوان (مقدمة الكشــف عــن نماذج ماركوف المخفية) ، إذ إنه من خلال مثال بسيط عن درجــات الحــرارة ، أعطى فكرة موجزة حول مفهوم نماذج ماركوف المخفية وحالاتها المخفية ومسائلها الأساسية وطرائق حلها فضلاً عن طريقــة البرمجــة الحركيــة وكيفيــة تقيــيس (Scaling) نموذج ماركوف المخفي (Stamp ,2003) .

(2) هدف الرسالة :

تهدف هذه الرســالة إلى عرض النظريــة الأســاسية لــ(HMMs) والتي تعود لــ(L.E. Baum) وكذلك التعرف على أهم المسائل المتعلقة بها فضلاً عــن وصف بعض التطبيقات حول النظرية ، وتطبيقها على التمييــز بــين الحــروف الصحيحة والعلة في اللغة الإنكليزية .

(3) العملية التصادفية (Stochastic Process) :

إن أية ظاهرة حقيقية تجري في حيز المعلمة المدروسة (كالزمن مثلاً) هي عملية تصادفية إذا كانت حالات تلك الظاهرة في أي جزء من حيزها (في أي وقت مثلاً) تمثل نتائج تجربة عشوائية تخضع لقوانين الاحتمالات ، هذا وتعرف العملية التصادفية أيضاً بالعملية العشــوائية (Random Process) أو عمليــة الفرصــة (Chance Process) .

تعريف: العملية التصادفية $\{X(t) , t \in T\}$ هي العملية المــكونَة مــن خــــلال قيــم مختلفة تأخذها متغيرات عشوائية معينة ، معرَّفة على تجربة ما ،عنــد قــيم مختلفة من فضاء المعلمة المدروسة (عند أزمان مختلفة مثلاً) حيث إن (T) تمثــل مجمــوعة المعلــمات وتشــير إلى فضاء المعلمة (Parameter Space) .

(4) فضاء الحالة (The State Space) :

تعريف : وهو مجموعة من القيم المتعلقة بـــ$X(t)$.

ويكون فضاء الحالة متقطعاً (Discrete) إذا احتوى على نقاط منتهية أو نقاط غير منتهية قابلة للعد ، أما في الحالات الأُخرى فهو مستمر(Continuous) .

(Cox & Miller , 1965 , P.14) .

(5) سلاسل ماركوف (Markov Chains) :

معظم النظم الشائعة الاستخدام في التطبيقات العملية تتمتع بخاصية كــون أن حالة الظاهرة قيد التحليل في الزمن الحاضر والزمن الماضي هي التي تحدد حالتها في المستقبل ، ولربما قد تدخل في ذلك عوامل أُخــرى وذلــك حســب المســألة المدروسة .

أما ما يخصنا فهو تلك النظم والظواهر التي تتميز بخاصية أنه إذا عُلمت حالة الظاهرة في الزمن الماضي ، فإنه سوف لن يكون هناك أي تأثير لحالات الظاهرة في الزمن الماضي على ما سيجري لها في المستقبل ، وتلك النظم التي تمتلك هذه الخاصية تُعرف بسلاسل مــاركوف ، وإن الخاصــية نفســها تُــدعى بالخاصــية الماركوفية (Markovian property) (الربيعي وعبد،2000 ،ص15) .

(6) المصفوفة الاحتمالية الانتقالية :

(The Transition Probability Matrix)

إن المصفوفة العشوائية (التصادفية) هي مصفوفة ذات عناصر عشــوائية(غيــر سالبة) منتهية أو غير منتهية بحيث إن مجموع عناصر كل صف يساوي (1) ، فإذا كانت المصفوفتان $A= [a_{ij}]$ و$B= [b_{ij}]$ تصادفيتين فإن المصفوفة الناتجــة عــن حاصــــل ضــــــربهما هــــــي أيضــــــاً مصــــــفوفة تصــــــادفية (Srinvasan & Mehata ,1976 , P.35) .

أما $p(i, j)$ فهي تمثل احتمالية الانتقال من الحالة (i) في الخطوة (n) إلــى الحالة (j) في الخطوة (n+1) حيث يتم وضع العناصر $p(i , j)$ (لكــل $i , j \in E$) على شكل مصفوفة مربعة مثلاً (P) وبافتراض أن (E) هي $\{0, 1 , 2 , \ldots\}$ ، فإنه يمكن كتابة المصفوفة (P) بالشكل الآتي :

$$P = \begin{bmatrix} p(0,0) & p(0,1) & p(0,2) & \cdots & p(0,N) \\ p(1,0) & p(1,1) & p(1,2) & \cdots & p(1,N) \\ p(2,0) & p(2,1) & p(2,2) & \cdots & p(2,N) \\ \vdots & \vdots & \vdots & & \ddots \\ p(N,0) & p(N,1) & p(N,2) & & p(N,N) \end{bmatrix}$$

تعريف : يُقال للمصفوفة المربعة (P) (المعرَّفة أعلاه) بأنها مصـــفوفة مـــاركوف (أو مصفوفة الانتقال) المُمثلة للعملية $\{X_n, n \in N\}$ ،على فضاء الحـــالة (E) ، إذا تحقق الشرطان الآتيان **(الربيعي وعبد،2000 ،ص17)** :

1. لكل قيم $(i , j \in E)$ تكون $p(i , j) \geq 0$.
2. لكل قيم $(i \in E)$ يكون $\sum_{j \in E} p(i, j) = 1$.

(7) نماذج ماركوف المخفية :

نماذج ماركوف المخفية هي مجموعة منتهية من الحالات كل حالة تقترن بتوزيع احتمالي ، إما الانتقالات ما بين الحالات فتُحدد بوســاطة مجموعــة مــن الاحتمالات وتُسمى الاحتمالات الانتقالية . بشكل عام ، تتولد الحالــة الناتجــة (أي المشاهدة) طبقاً لتوزيع الاحتمالية المقترنة ، حيث توجد احتمالية ناتجــة فقــط ولا توجد حالة ظاهرة يمكن أن تُشاهد ، ولهذا فإن الحالات تكون مخفية ؛ هــذا هــو معنى نماذج ماركوف المخفية بشكل عام **(Rabiner , 1989)** .

وهناك تعاريف لمفهوم نماذج ماركوف المخفية منها :

تعريف: وهي تمثل متسلسلات تصادفية (كما في سلاسل مــاركوف) حيـــث إن الحـــالات لا تُــشاهد بشكل مباشر ولكــنها تقترن بدالة كثافة احتمالية (pdf) ، وبشكل أوضـــح فــإن متسلســلة الحالــة $Q =\{q_n\}$ (أي متسلسلة الحالات المخفية) لا يمكن مشاهدتــها ولــكن المشــاهدات (O) تعتمــد على (Q) ،أي $P(O|Q)$ (Ellis , 2001) .

تُعدّ نماذج ماركوف المخفية إحدى الوسائل المفيدة لدراسة النماذج الاحتمالية في السلاسل الزمنية، إذ إن معلومات (HMMs) حول الماضي تُنقل على شكل متغير متقطع منفرد (وهي تمثل الحالة المخفية) .

إن نماذج ماركوف المخفية هي نماذج إحصائية تُستخدم لنمذجة البيانات حيث إنها تُستخدم وبنجاح في العديد من المجالات منها تمييز وفهم الكلام وسيطرة الرجل الآلي (Zhai , 2003) .

في الآونة الأخيرة ، حازت (HMMs) على اهتمام كبير في مجتمع الـ(Bioinformatic) (أي مجتمع المعلوماتية الحياتية) ، إذ إنها تُستخدم في حل المسائل الخاصة بتنبؤ التركيب الثانوي البروتيني (Protein Secondary Structure Prediction) وإيجاد الجينة ، وتصنيف عائلة البروتين وتشكيل خرائط ارتباط الجينات لعلم الوراثة وخرائط فيزياوية (Parker ,1999) .

(8) عناصر نماذج ماركوف المخفية (Elements of HMMs) :

ويمكن تلخيص عناصر(HMMs) بالآتي :

T = طول متسلسلة المشاهدات .

N = عدد الحالات المخفية في النموذج .

M = عدد رموز المشاهدات .

$Q = \{q_0 , q_1 , \ldots , q_{N-1}\}$ = الحالات لعملية ماركوف (أي الحالات المخفية) .

$V = \{0 , 1 , \ldots , M-1\}$ = مجموعة رموز المشاهدات .

A = مصفوفة احتمالات انتقال الحالة .

B = مصفوفة احتمالية رابطة بين الحالات المخفية والمشاهدات .

π = متجه توزيع الحالة الابتدائية .

$O = \{O_0 , O_1 , \ldots , O_{T-1}\}$ = متسلسلة المشاهدات .

وللسهولة يمكن استخدام العبارة المركبة $\lambda = (A , B , \pi)$ للإشارة إلى مجموعة معلمات النموذج(Rabiner , 1989) .

(9) أنواع نماذج ماركوف المخفية (Types of HMMs) :

يمكن تصنيف (HMMs) إلى نوعين وذلك حسب الانتقالات بين حالاتها وكما يأتي :

1. النموذج الثبوتي (Ergodic Model) : وهو النموذج الذي تكون فيهِ كل الحالات انتقالية.
2. نموذج الأيسر–الأيمن (Left-Right Model) : وهو النموذج الذي تكون فيهِ بعض الحالات انتقالية ، بحيث إن :

$$a_{ij} = 0 \qquad , \forall \quad j < i$$

أي أنها مصفوفة مثلثية عُليا ، وهذا النوع من النماذج يُستخدم وبشكل واسع في نمذجة متسلسلات الإشارات (Rabiner , 1989) .

(10) المسائل الأساسية لـنماذج ماركوف المخفية :
(The Basic Problems for HMMs)

المسألة (1) :

إذا كان المعطى هو النموذج $\lambda = (A , B , \pi)$ ومتسلسلة المشاهدات $O = O_1O_2\ldots O_T$ ، أوجد $P(O|\lambda)$ ، أي تحديد الإمكان لمتسلسلة المشاهدات (O) عندما يكون المعطى هو النموذج .

المسألة (2) :

إذا كان المعطى هو النموذج $\lambda = (A , B , \pi)$ ومتسلسلة المشاهدات $O = O_1O_2\ldots O_T$ ، أوجد متسلسلة الحالة المثلى (The Optimal State Sequence) لعملية ماركوف المخفية ، والتي تمثل أفضل توضيح للمشاهدات .

المسألة (3) :

إذا كـــان المعطـــى هـــو متسلســـلة المشـــــاهدات $O = O_1 \ O_2 \ \ldots O_T$ والأبـــعاد(N) و(M)، أوجد النموذج $\lambda = (A , B , \pi)$ الذي تكون فيهِ $P(O \mid \lambda)$ عُظمى ، أي إعادة تقدير معلمات النمـــوذج $\lambda = (A , B, \pi)$ لتعظـــيم $P(O \mid \lambda)$ (Stamp,2003) .

(11) حلول المسائل الأساسية لــنماذج ماركوف المخفية :

(Solutions to the Basic Problems of HMMs)

(11.1) حل المسألة (1) :

إن الطريقة المباشرة لحسابها تكون من خلال سرد متسلسلة حــــــالة ذات طـــــول (T)(حيث إن (T) تمثل عدد المشاهدات) ولنتأمل المتسلسلة الثابتة للحالة :

$$Q = q_1 \, q_2 \ldots q_T \qquad \ldots(1)$$

حيث إن (q_1) تمثل الحالة الابتدائية ، أما الاحتمالية لمتسلســـلة المشـــاهدات (O) بالنسبة لمتسلسلة الحالة في المعادلة (1) فهي (Rabiner , 1989) :

$$P(O \mid Q , \lambda) = \prod_{t=1}^{T} P(O_t \mid q_t , \lambda) \qquad \ldots(2\,a)$$

وذلك بافتراض الاستقلال الإحصائي للمشاهدات ، أي إن :

$$P(O \mid Q,\lambda) = P(O_1 \mid q_1 , \lambda) * P(O_2 \mid q_2 , \lambda) * \ldots * P(O_T \mid q_T,\lambda)$$

$$= \prod_{t=1}^{T} P(O_t \mid q_t , \lambda)$$

$$= \prod_{t=1}^{T} b\,q_t(O_t) \qquad \ldots(2\,b)$$

وأن الاحتمالية لمتسلسلة الحالة (Q) هي :

$$P(Q|\lambda) = P(q_1, q_2, \ldots, q_T|\lambda)$$
$$= P(q_1; \lambda) \prod_{t=2}^{T} P(q_t|q_{t-1}; \lambda)$$
$$= \pi q_1 \prod_{t=2}^{T} a\, q_{t-1}, q_t \qquad \ldots(3)$$

وذلك حسب الخاصية الماركوفية (Ellis , 2001) حيث إن (π) تشير الى احتمالية الحالة الابتدائية للحالة (q_1) .

أما الاحتمالية المشتركة لـــ(O) و(Q) ، أي احتمالية ورود (O) و(Q) بالوقت نفسه ، فهي تمثل النتيجة للمعادلتين (2.5 b) و(2.6) ، أي إن :

$$P(O, Q|\lambda) = P(O|Q, \lambda) * P(Q|\lambda) \qquad \ldots(4)$$

ومن هنا يمكن الحصول على الاحتمالية لـــ(O) عندما يكون المعطى هو النموذج (λ) وذلك عن طريق جمع هذهِ الاحتمالية المشتركة لكل متسلسلات الحالة (q) ، حيث إن :

$$P(O|\lambda) = \sum_{all\,Q} P(O|Q, \lambda) * P(Q|\lambda) \qquad \ldots(5)$$

$$= \sum_{q_1, q_2, \ldots, q_T} \pi_{q1} * b_{q1}(O_1) * a_{q1q2} * b_{q2}(O_2) * \ldots * a_{qT-1qT} * b_{qT}(O_T) \qquad \ldots(6)$$

إن هذهِ الحسابات تكون غير عملية وإن كانت قيم (N) و(T) صغيرة ، فمـــثلاً إذا كانت (N = 5) (وهي تمثل عدد الحالات المخفية) وكانت (T = 100) (وهي تمثل عدد المشاهدات) ، فيجب القيام بـــ $2(100) * 5^{100} \approx 10^{72}$ عملية حسابية ، ومن ذلك يتضح بأن هذهِ الحسابات تتطلب جهداً كبيراً لحل المسألة (1) ، ولذلك استخدم البـــاحثون الخوارزميـــات الأماميـــة – الخلفيـــة (Forward–Backward Algorithms) لحل تلك المسألة (Rabiner , 1989) ، وفيما يأتي توضيح لهـــذهِ الخوارزميات :

أولاً :ـ الخوارزميات الأمامية :

تُعرَّف الاحتمالية الأمامية $\alpha_t(i)$ على أنها الاحتمالية لمتسلسلة مشاهدات جزئية والحالة $(q_t=S_i)$ عندما يكون المعطى هو النموذج (λ) ، أي :

$$\alpha_t(i) = P(O_1 O_2 \dots O_t, q_t = S_i \mid \lambda) \qquad \dots(7)$$.

حيث إن الاحتمالية المطلوب إيجادها هي :

$$P(O \mid \lambda) = \sum_{i=0}^{N-1} \alpha_{T-1}(i)$$

وهذهِ الاحتمالية يمكن إيجادها بشكل متعاقب (Recursion) وكالآتي :

1. البداية (Initialization) :

$$\alpha_0(i) = \pi_i b_i(O_0) \qquad , i = 0, \dots, N-1 \qquad \dots(8)$$

2. التعاقب (Recursion) :

$$\alpha_t(j) = \left[\sum_{i=0}^{N-1} \alpha_{t-1}(i)\, a_{ij}\right] * b_j(O_t) \qquad , t = 1, \dots, T-1 \qquad \dots(9)$$

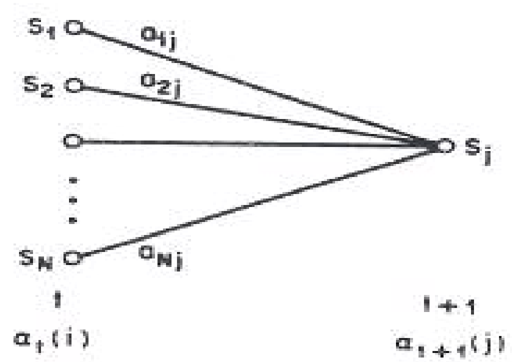
$$, j = 0, \dots, N-1$$

3. النهاية (Termination) :

$$P(O \mid \lambda) = \sum_{i=0}^{N-1} \alpha_{T-1}(i) \qquad \dots(10)$$

ومن هنا تُحسب $P[O_1 O_2 \dots O_T]$ (Parker ,1999) .

ويمكن توضيح الخوارزميات الأمامية من خلال الشكل **(1)** .



الشكل (1) : سلسلة العمليات المطلوبة لحساب الاحتمالية الأمامية α_{t+1}

إن حسابات الــ(α) تتطلب $N(N+1)(T-1)+N$ عملية ضرب و$N(N-1)(T-1)$ عملية جمع وهي أفضل من الحسابات المباشرة والتي تتطلـــب ($2T * N^T$) عمليـــة حسابية (Stamp ,2003) و(Rabiner ,1989) .

ثانياً :ـ الخوارزميات الخلفية :

تُعرَّف الاحتمالية الخلفية على أنها الاحتمالية لمتسلسلة المشاهدات الجزئية $O_{t+1},O_{t+2},\ldots,O_T$ عندما يكون المعطى هو الحالـــة ($q_t = S_i$) والنمـــوذج (λ) وكما يأتي :

$$\beta_t(i) = P(O_{t+1}\, O_{t+2} \ldots O_T \mid q_t = S_i, \lambda) \qquad \ldots(11)$$

وهـــذهِ الاحتماليـــة يمكـــن إيجادهـــا أيضـــاً بشـــكل متعاقـــب(Recursion) وكالآتي(Stamp,2003) :

1. البداية :

$$\beta_{T-1}(i) = 1 \qquad , i = 0, \ldots, N-1 \qquad \ldots(12)$$

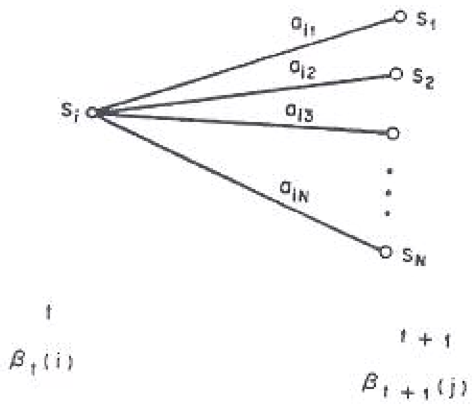
2. التعاقب :

$$\beta_t(i) = \sum_{j=0}^{N-1} a_{ij} * b_j(O_{t+1}) * \beta_{t+1}(j) \quad , t = 0, \ldots, T\text{-}2 \quad , i = 0, \ldots, N\text{-}1 \qquad \ldots(13)$$

3. النهاية :

$$P(O \mid \lambda) = \sum_{i=0}^{N-1} \alpha_t(i) * \beta_t(i) \text{ , for any } t \qquad \ldots(14)$$

والشكل (2) يوضح العمليات المطلوبة لحساب الاحتمالية الخلفية .



الشكل(2) : سلسلة العمليات المطلوبة لحساب الاحتمالية الخلفية $\beta_t(i)$

(1.21) حل المسألة (2) :

هناك بعض الطرائق الممكنة لحل المسألة (2) ، أي لإيجاد متسلسلة الحالة المُثلى المقترنة بمتسلسلة المشاهدات المعطاة منها :

خوارزمية (Viterbi) :

هي طريقة البرمجة الحركية المُستخدمة لحساب مسار الحالـة الانتقاليـة الأكثر احتمالاً عندما يكون المعطى هو متسلسلة المشاهدات للرموز، وهي مشابهة للخوارزمية الأمامية ما عدا كون استخدام الــ" أكبر " " Max " بدلاً من الـ" مجموع " " $\sum$ " (Zhai , 2003) .

ولإيجاد أفضل متسلسلة حالـــة وحيـــدة $Q= [q_0\ q_1\ \dots\ q_{T-1}]$ بالنســـبة لمتسلسلة المشاهدات المعطاة $O = [O_0\, O_1 \dots O_{T-1}]$ يجب تعريف المقدار :

$$\delta_t(i) = \max_{q_0 \dots q_{t-2}} P[q_0\, q_1 \dots q_{t-1} = S_i ,\ O_0\, O_1\ \dots\ O_{t-1} | \lambda] \quad \dots(15)$$

إذ إن $\delta_t\ (i)$ تمثل أعلى احتمالية على طول المسار الوحيـد خـــلال الـــزمن (t) والمحسوبة بالنسبة لمشاهدات (t) الأولى والمنتهية في الحالة (S_i) ، فضلاً عـــن استخدام $\psi_t\ (i)$ والتي تساعد في حفظ تتابع الأثر (Keep Track) للمسار الفعلي ، أي إن $\psi_t\ (i)$ ستوضح أية حالة خلال الـــزمن (t–1) تقـــود لاحتماليـــة $\delta_t\ (i)$ الأعلى خلال الزمن (t) (Parker ,1999) .

أما خطوات خوارزمية (Viterbi) فهي كالآتي :

1. البداية :

$$\delta_0(i) = \pi_i * b_i(O_0)\ ,\ i = 0, \dots, N-1 \qquad \dots(16\ a)$$

$$\psi_0(i) = 0 \qquad \dots(16\ b)$$

2. التعاقب :

$$\delta_t(j) = \max_{0\le i\le N-1} [\delta_{t-1}(i) * a_{ij}] * b_j(O_t) \quad ,\ t = 1, \dots, T-1 \quad ,\ j = 0, \dots, N-1 \qquad \dots(17\ a)$$

$$\psi_t(j) = \arg\max_{0\le i\le N-1} [\delta_{t-1}(i) * a_{ij}]\ ,\ t = 1, \dots, T-1 \quad ,\ j = 0, \dots, N-1 \qquad \dots(17\ b)$$

3. النهاية :

$$P^* = \max_{0\le i\le N-1} [\delta_{T-1}(i)] \qquad \dots(18\ a)$$

$$q^*_{T-1} = \arg\max_{0\le i\le N-1} [\delta_{T-1}(i)] \qquad \dots(18\ b)$$

4. التتابع المعاكس (Back Tracking) :

$$Q^* = \{q^*_0 , \dots , q^*_{T-1}\}$$

$$q^*_t = \psi_{t+1}(q^*_{t+1}) \qquad ,\ t = T-2\ ,\ T-3\ , \dots ,\ 0 \qquad \dots(19)$$

ثم تُقرأ المتسلسلة الأفضـــل للحالات من مُتجهات الـــ(ψ_t) .

(11.3) حل المسألة (3) :

المطلوب في المسألة (3) هو إعادة تقدير معلمات النموذج (A , B , π) وذلك لتعظيم الاحتمالية لمتسلسلة المشاهدات عندما يكون المعطى هو النموذج ويمكن حل هذه المسألة باستخدام خوارزمية (Baum–Welch) (Rabiner , 1989) .

خوارزمية (Baum–Welch) :

هذهِ الخوارزمية تُناقش خطوات التكرار بالاعتماد وبشكل ابتدائي على العمل الكلاسيكي للعالم (L.E. Baum) وزملائهِ وذلـــك لاختيـــار معلمـــات النمـــوذج . ولوصف خطوات إعادة تقديـــر (Re–estimate) معلمات (HMM) ، يتم تعريف المتغيرين الآتيين :

المتغير الأول هو :

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=0}^{N-1}\alpha_t(i)\beta_t(i)}$$

أما المتغير الثاني فهو :

$$\xi_t(i , j) = \frac{\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{P(O\,|\,\lambda)}$$

$$= \frac{\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)} \quad ، t = 0, \ldots, T\text{-}2 \qquad \ldots(20)$$

وباستخدام الصيغ أعلاه ، يمكـــن إدراج الصيغ الآتية وذلك لإعـــادة تقدير معلمات (HMM) (A , B , π) (Stamp ,2003) :

$$\bar{\pi}_i = \gamma_o(i) \qquad , i = 0, \ldots, N\text{-}1 \qquad \ldots(21\ a)$$

$$\bar{a}_{ij} = \frac{\sum_{t=0}^{T-2} \xi_t(i, j)}{\sum_{t=0}^{T-2} \gamma_t(i)} \qquad , i , j = 0, \ldots, N\text{-}1 \qquad \ldots(21\ b)$$

$$\bar{b}_j(k) = \frac{\sum_{\substack{t\in(0,1,\ldots,T-2) \\ O_t = vk}} \gamma_t(j)}{\sum_{t=0}^{T-2} \gamma_t(j)} \qquad , j = 0, \ldots, N\text{-}1 \qquad \ldots(21\ c)$$

بالاعتماد على الخطوات السابقة وباستخدام (وبشكل متكرر) $\bar{\lambda} = (\bar{\pi}, \bar{B}, \bar{A})$ بدلاً من $\lambda = (A , B , \pi)$ وتكــرار حسابات إعادة التقــدير فســوف تُعـــدّل الاحتمالية لــ(O) المُشاهدة من النموذج إلى أن يتم التوصل إلـــى بعــض النقــاط المُحددة ، والنتيجة النهائية لخطوات إعادة التقدير هذهِ تُعرف بــ(تقــدير الإمكــان الأعظم لــ(HMM)) .

إن الجانب المهم من خطوات إعــادة التقدير هو شــرط (أو قيــد) التصـــــادفية لمعلمات الـــ(HMM)، وبالذات المتحققة في كل تكرار،أي إن (Rabiner, 1989)

$$\sum_{i=0}^{N-1} \bar{\pi}_i = 1$$

$$\sum_{j=0}^{N-1} \bar{a}_{ij} = 1 \qquad , i = 0, \ldots, N\text{-}1$$

$$\sum_{k=0}^{M-1} \bar{b}_j(k) = 1 \qquad , j = 0, \ldots, N\text{-}1$$

إن عملية إعــادة التقدير التي تمثل عملية التكرار بدأت من اختيار أفضـــل قيم تخمينية (أو تقديرية) لــ$(A , B , \pi) = \lambda$ ، وإذا لم تتوافر هذه القيم فســوف يتم اختيار قيم عشوائية بحيث إن :

$$\pi_i \approx 1/N \qquad a_{ij} \approx 1/N \qquad b_j(k) \approx 1/M$$

مع التأكيد على أن تكون (A , B , π) دائماً ذات صف تصادفي ، وأن لا تكون قيم عناصر المصفوفة متساوية .

ويمكن تلخيص العملية بالخطوات الآتية :

1. البداية ، اختيار قيم عشوائية ابتدائية للنموذج $\lambda = (A , B , \pi)$.
2. حساب $\alpha_t(i)$ ، $\beta_t(i)$ ، $\xi_t(i,j)$ ، $\gamma_t(i)$.
3. إعادة تقدير النموذج $\lambda = (A , B , \pi)$.
4. إذا كانت $P(O|\lambda)$ تزايدية ، فيجب الذهاب إلى (2) .

وبالطبع يجب التوقف إذا لم تكن $P(O|\lambda)$ تزايدية (Stamp ,2003) .

(12) تقييس نماذج ماركوف المخفية(Scaling HMMs) :

يمكن تنفيذ الصيغ المُعطاة في المباحث السابقة كما هي ولكنها تكون مفيدة في حالة كون المتسلسلات قصيرة جداً ، أما عندما تكون المتسلسلة طويلة فإن عدة مقادير تصغر وبشكل حاد ولحل هذهِ المشكلة هناك طريقتان :

الأولى ، العمل على مجال اللوغاريتم ، أي أخذ اللوغاريتم الــذي يُقــرب النـــاتج للكميات الصغيرة .

الثانية ،طريقة الـــ(Scaling) أي التقييس (وهي الطريقة المفضلة) والقصد منهـــا تقييس $\alpha_t(i)$ و$\beta_t(i)$ ،حيث يُرمز لهما بـــ$\hat{\alpha}_t(i)$ و $\hat{\beta}_t(i)$ (على التوالي) ، على أن تكون $\hat{\alpha}_t(i)$ مُتناسبة مع $\alpha_t(i)$ ويكون المجموع يساوي (1) لكل الحالات الممكنة (أي لجميع قيم i) وكالآتي (Zhai , 2003) و(Stamp ,2003) :

$$\sum_{i=0}^{N-1} \hat{\alpha}_t(i) = 1$$

هذا وأن :

$$\hat{\alpha}_t(i) = \prod_{k=0}^{t} \eta_k \ \alpha_t(i) \quad , i = 0,\ldots, N-1 \ ; t = 0, \ldots, T-1 \qquad \ldots(22)$$

وكذلك فإن :

$$\prod_{k=0}^{T-1} \eta_k = \frac{1}{\sum_{i=0}^{N-1} \alpha_{T-1}(i)} = \frac{1}{P(O|\lambda)}$$

أو :

$$\log[P(O|\lambda)] = -\sum_{k=0}^{T-1} \log(\eta_k) \qquad \ldots(23)$$

حيث يمكن حساب لوغاريتم(P) وليس (P) ، هذا وأن التقييس سيمر بعدة خطوات كما هي مبينة أدناه .

أولاً: خوارزمية تقييس الاحتمالية الأمامية (α) :

(The Scaling Forward Algorithm)

يتم حساب ($\hat{\alpha}$) بنفس خوارزمية (α) إجراء بعض التغييرات عند كل خطوة وكالآتي (Zhai , 2003) :

$$1.\ \hat{\alpha}_o(i) = \frac{\pi_i b_i(O_0)}{\sum_{k=0}^{N-1} \pi_k b_i(O_0)} \quad ، i = 0, \ldots, N-1 \qquad \ldots(24\ a)$$

$$2.\ \hat{\alpha}_t(i) = \frac{b_i(O_t) \sum_{j=0}^{N-1} \hat{\alpha}_{t-1}(j) a_{ji}}{\sum_{k=0}^{N-1} b_k(O_t) \sum_{j=0}^{N-1} \hat{\alpha}_{t-1}(j) a_{jk}} \quad , i = 0, \ldots, N-1 \qquad \ldots(24\ b)$$

$$, t = 1, \ldots, T-1$$

ثانياً : خوارزمية تقييس الاحتمالية الخلفية (β):

(The Scaling Backward Algorithm)

يمكن تقييس الاحتمالية الخلفية (β) بالطريقة المستخدمة نفسها لتقييس الاحتمالية الأمامية (α) ذلك إن :

$$\hat{\beta}_t(i) = \beta_t(i) \prod_{k=t+1}^{T-1} \eta_k \qquad \ldots(25)$$

حيث إن $\hat{\beta}_T(i) = \beta_T(i)$ ، وأن :

$$\hat{\alpha}_t(i)\hat{\beta}_t(i)=\prod_{k=0}^{T-1}\eta_k\alpha_t(i)\beta_t(i)=\frac{\alpha_t(i)\beta_t(i)}{P(O|\lambda)}$$

حيث يتم حساب الاحتمالية الخلفية ($\hat{\beta}$) بطريقة حساب الاحتمالية الخلفية (β) نفسها مع إدخال (η) (المحسوبة سابقاً من خوارزمية تقييس الاحتمالية الأمامية (α)) في القانون وكالآتي (Zhai , 2003) :

1. $\hat{\beta}_{T-1}(i)=\beta_{T-1}(i)=1$, i = 0, …, N-1 …(26 a)

2. $\hat{\beta}_t(i)=\eta_{t+1}\sum_{j=0}^{N-1}\hat{\beta}_{t+1}(j)\,a_{ij}\,b_j(O_{t+1})$ ، t = 0, …, T-1 …(26 b)

ثالثاً: صيغ التحديث باستخدام($\hat{\alpha}$) و($\hat{\beta}$) :

(The Updating Formulas Using Scaled α ‘s and β‘s)

يمكن حساب الــ(γ) باستخدام الــ($\hat{\alpha}$) والــ($\hat{\beta}$) وكالآتي :

$$\gamma_t(i)=\frac{\hat{\alpha}_t(i)\hat{\beta}_t(i)}{\sum_{j=0}^{N-1}\hat{\alpha}_t(j)\hat{\beta}_t(j)}\quad , \; i = 0, \ldots, N-1 \; ; \; t = 0, \ldots, T-2 \quad \ldots(27)$$

وحساب (ξ) بصيغ مختلفة وكالآتي :

$$\xi_t(i,j)=\frac{\hat{\alpha}_t(i)a_{ij}b_j(O_{t+1})\,\eta_{t+1}\hat{\beta}_{t+1}(j)}{\sum_{j=0}^{N-1}\hat{\alpha}_t(j)\hat{\beta}_t(j)}\quad \ldots(28)$$

أو :

$$\xi_t(i,j)=\frac{\gamma_t(i)a_{ij}b_j(O_{t+1})\,\eta_{t+1}\hat{\beta}_{t+1}(j)}{\hat{\beta}_t(i)}\quad \ldots(29)$$

وباستخدام (γ) و(ξ) المحسوبتين أعلاه ، يمكن حساب (a , b , π) المعدلة [والتي يُرمز لها بــ($\bar{a}$, $\bar{b}$, $\bar{\pi}$)] وكالآتي (Zhai , 2003) و(Stamp ,2003) :

$\bar{\pi}_i=\gamma_o(i)$, i = 0, …, N-1 …(30 a)

$$\bar{a}_{ij} = \frac{\sum_{t=0}^{T-2} \xi_t(i,j)}{\sum_{t=0}^{T-2} \gamma_t(i)} \qquad , i , j = 0, \ldots, N-1 \qquad \ldots(30\ b)$$

$$\bar{b}_j(k) = \frac{\sum_{\substack{t\in(0,1,\ldots,T-2)\\ O_t=vk}} \gamma_t(j)}{\sum_{t=0}^{T-2} \gamma_t(j)} \qquad , j = 0, \ldots, N-1 \qquad \ldots(30\ c)$$

رابعاً: خوارزمية تقييس الاحتمالية (δ) :

أما عند استخدام خوارزمية (Viterbi) لإعطاء الإمكان لمتسلسلة الحالة ، فليس من الضروري عمل (Scaling) إذا استخدمت اللوغاريتمات بالطريقة الآتية، حيث يُعرَّف المقدار $\delta_t(i)$ كالآتي :

$$\delta_t(i) = \max_{q_1,q_2,\ldots,q_t} \{\log P[q_1 q_2 \ldots q_t, O_1 O_2 \ldots O_t|\lambda]\} \qquad \ldots(31)$$

وهذهِ الاحتمالية يمكن إيجادها بشكل تعاقبي وكالآتي :

1. البداية :

$$\delta_0(i) = \log(\pi_i) + \log[b_i(O_0)] \qquad , i = 0, \ldots, N-1 \qquad \ldots(32\ a)$$

$$\psi_0(i) = 0 \qquad \ldots(32\ b)$$

2. التعاقب :

$$\delta_t(j) = \max_{0\le i\le N-1}[\delta_{t-1}(i) + \log(a_{ij})] + \log[b_j(O_t)], t=1,\ldots,T-1 \qquad (33\ a)$$

$$, j=0, \ldots, N-1$$

$$\psi_t(j) = \arg\max_{0\le i\le N-1} [\delta_{t-1}(i) + \log(a_{ij})] \qquad \ldots(33\ b)$$

3. النهاية :

$$\log P^* = \max_{0\le i\le N-1}[\delta_{T-1}(i)]$$
…(34 a)

$$q^*_{T-1} = \arg\max_{0\le i\le N-1}[\delta_{T-1}(i)] \qquad \ldots(34\ b)$$

4.التتابع المعاكس :

$$Q^* = \{q^*_0,\ldots,q^*_{T-1}\}$$
$$q^*_t = \psi_{t+1}(q^*_{t+1}) \qquad , t = T-2, T-3, \ldots, 0 \qquad \ldots(35)$$

(Rabiner ,1989) و(Stamp ,2003) .

(31)الجانب التطبيقي :

تم في هذا البحث تطبيق نماذج ماركوف المخفية للتمييز بين الحروف الصحيحة وحروف العلة في نصوص اللغة الإنكليزية .

(13.1) خطوات العمل :

الخطوة الأولى :

ترتيب الحروف الإنكليزية بشكل متسلسل حيث يقابل كل حرف رقماً يمثله، وكذلك وضع رقم للفراغ (Space) بين كلمة وأخرى .

الخطوة الثانية :

تؤخذ سلسلة من المشاهدات التي هي حروف مرتبة حسب النص الذي أُخذت منه ، وتكون كما يأتي :

إذا كان الحرف في الكلمة هو (a) فيرمز له بـ(0) وإذا كان الحرف هو (b) فيرمز له بـ(1) وهكذا لبقية الحروف .

الخطوة الثالثة :

هي عملية اختيار قيم ابتدائية للمعلمات،ليست هناك طريقة مباشرة لاختيار التقديرات الابتدائية،ولعمل ذلك تم اختيار القيم بشكل عشوائي أو تخميني .

الخطوة الرابعة :

هي عملية حساب المتغيرات الأمامية (أي حساب قيم α وقيم α المعدلة) .

الخطوة الخامسة :

هي حساب المتغير (η) ويكون بالاعتماد على قيم كل من (α) و $(\hat{\alpha})$ فضلاً عن إيجاد $\log P(O|\lambda)$.

الخطوة السادسة :

وهي حساب المتغيرات الخلفية (أي حساب β) وبسبب طول متسلسلة المشاهدات أيضاً فسوف يتم تقييس قيم (β) .

الخطوة السابعة :

وتتمثل في حساب كل من $\gamma_t(i)$ و $\xi_t(i, j)$ وذلك حسب صيغ التحديث باستخدام $(\hat{\alpha})$ و $(\hat{\beta})$.

الخطوة الثامنة :

إعادة حساب المعلمات $\lambda = (A, B, \pi)$ لتعظيم الاحتمالية لمتسلسلة المشاهدات إذا كان المعطى هو النموذج مع التأكيد على أن تكون المصفوفات تصادفية ، ومن هذه المصفوفات الناتجة يُعاد حساب $\log P(O|\lambda)$ فإذا كانت تزايدية تُعاد الخطوات السابقة (من الخطوة الرابعة) وذلك باستخدام المصفوفات الجديدة الناتجة لــ(A) ، (B) و(π) كمصفوفات داخلة .

(13.2) النتائج :

بعد حساب المتغيرات المطلوبة وإجراء التكرار الابتدائي لوحظ بأن :

$$\log P(O|\lambda) = -634.433173$$

وبعد التكرار $(q = 284)$:

$$\log P(O|\lambda) = -512.467179$$

وهذا يؤكد أن النموذج قد تحسن وبشكل واضح ، إذ إن النموذج $\lambda = (A, B, \pi)$ بعد التكرار (284) تحول إلى :

$$\overline{\pi} = [1.000000 \qquad 0.000000]$$

$$\overline{A} = \begin{bmatrix} 0.452615 & 0.547385 \\ 0.960543 & 0.039457 \end{bmatrix}$$

فضلاً عن مبدلة المصفوفة $(\overline{B})$ الموضحة في الجدول .

إن الجزء الأكثر أهمية في هذه النتيجة هو المصفوفة (B) ، وبشكل واضح وبدون أي افتراض حول الحالتين المخفيتين فإن المصــفوفة (B) تُظهر بأن هناك حــالة واحدة مخفية تحتوي على حروف العلة في حين إنّ الحالة المخفية الأخرى تحتوي على الحروف الصحيحة ، ذلك أن العمود الأول في الجدول (1) قد أعطى أعلى قيم لحروف العلة (وهي a , e , i , o وu) أما بقية الحروف (أي الحروف الصحيحة) فقد كانت صفرية أو مقاربة للصفر، لذا فإن العمود الأول يمثــل نســبة احتماليــة حروف العلة ، أما في العمود الثاني فقد كانت قيم حروف العلة مساوية للصفر، أي أن العمود الثاني يمثل نسبة احتمالية الحروف الصحيحة ، مع ملاحظة أن الحرف (y) يُعامل أحيانا كحرف علة .

(14)الاستنتاجات والتوصيات :

تم التوصل من خلال هذهِ الدراسة إلى ما يأتي :

في ما يخص الجانب التطبيقي لهذهِ الدراسة يُلاحظ أنهُ يمكــن معرفــة الحــروف الصحيحة وحروف العلة دون معرفة سابقة بقواعد اللغة الإنكليزية ، وقد تبين مــن الجدول (1) أن حروف العلة هي (a , e , i , o و u).

ويمكن التوصية بما يأتي :

1. من خلال مقابلة الباحثة لأساتذة اللغة الإنكليزية في كلية الآداب أوضــحوا ان هناك فرقاً بين النصوص العلمية والنصوص الأدبية لذا نوصي بتطبيق هذهِ النماذج للحالتين ومقارنة النتائج .

2. أكد المختصون أن هناك فرقاً بين حروف العلة وأصوات العلة والتــي هــي (12) صوتاً ، لذا نوصي بأخذ (12) حالة مخفية وملاحظة النتائج لها .

الجدول (1): قيم عناصر مبدلة المصفوفة) $\overline{B}$ (عند التكرار الأخير(284)

	القيم	
a	0.128153	0.000000
b	0.000000	0.018740
c	0.008775	0.019412
d	0.007727	0.068840
e	0.181550	0.000000
f	0.000000	0.031233
g	0.008478	0.004421
h	0.000000	0.187400
i	0.075755	0.000000
j	0.000000	0.003204
k	0.000000	0.018740
l	0.004951	0.048724
m	0.000000	0.062467
n	0.000000	0.181154
o	0.188995	0.000000
p	0.000000	0.031233
q	0.000000	0.001026
r	0.000000	0.118687
s	0.000516	0.100233
t	0.013922	0.021395
u	0.045919	0.000000
v	0.000000	0.024987
w	0.000000	0.054220
x	0.000000	0.000245
y	0.000665	0.001202
z	0.000000	0.001104
space	0.334594	0.001333

(15)المصادر :

1. الربيعي، فاضل محسن وعبد، صلاح حمـــزة (2000)." مقدمـــة فـــي العمليـــات التصادفية " ، دار الكتب – بغداد ، العراق .

2. Cox,D.R. and Miller, H.D. (1965). “ The Theory of Stochastic Processes ”, John Wiley & Sons , Inc. , New York , USA .

3. Ellis,D. (2001) “ Sequence Recognition ” , Computer, Speech, and Language, Vol. 1, no. 2, pp. 167-197

4. Parker,J. (1999). “ Profile Hidden Markov Models ” , Biophysical Journal , Vol. 17 , P. 1335 – 1348 .

5. Rabiner,L.R. (1989). “ A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition ” , Proceedings of IEEE , Vol. 77 , No. 2 , P. 257 – 286 .

6. Srinivasan,S.K. & Mehata, K.M. (1976). “ Stochastic Processes ” , Tata McGraw – Hill Publishing Company Limited , New Delhi .

7. Stamp,M. (2003). “ A Revealing Introduction to Hidden Markov Models ” , IEEE , Vol. 51 , No. 7 , P. 347 – 356 .

8. Zhai,C. (2003). “ A Brief Not on the Hidden Markov Models ” , IEEE Transactions on PAMI , Vol. PAMI –18 , No.3 , P.475–4 79. http://hanoi.cs.uiuc.edu/cs397/hmm.