AL- Rafidain University College

PISSN: (1681-6870); EISSN: (2790-2293)

مجلة كلية الرافدين الجامعة للعلوم

Available online at: https://www.jrucs.iq

JRUCS

Journal of AL-Rafidain University College for Sciences

الطرائق الذكية في تقدير معلمات انموذج ماركوف المخفي

أحمد عبد الزهرة دواي الكورجي	أ.د. حامد سعد نور الشمرتي
Ahmedalkorje0@gmail.com	Hamed_Saad@albayan.edu.iq
قسم الاحصاء، كلية الادارة و الاقتصاد، الجامعة المستنصرية، بغداد، العراق	كلية ادارة الاعمال، جامعة البيان، بغداد، العراق

المستخلص

نماذج ماركوف المخفية (HMMs) هي نماذج عشوائية تم تطبيقها في البدء كنماذج احصائية لتمييز الكلام والكتابة اليدوية بسبب قدرتها الكبيرة على التكيف مع المشكلة والبراعة في التعامل مع الاشارات المتسلسلة. ومع تطور التقنيات والادوات والطرائق الخاصة بتقدير معلمات انموذج ماركوف المخفي فأنه من الممكن ان تتوجه الانظار الى الطرائق الذكية لأهميتها واستعمالها الواسع لدى الباحثين، تم اخذ خوارزميتين وهما خوارزمية تحسين مستعمرة النمل Ant Colony Optimization(ACO) وخوارزمية تحسين مستعمرة النحل Bees colony optimization (BCO) وتم تطبيقهما على مجال هام في وقتنا الحالي وهو الامن السيبراني Cyber Security حيث تناولنا احد التهديدات التي تشكل خطرا جسيما وهو التهديد المستمر المتقدم Advanced Persistent Threat(APT). بينت النتائج المرونة في تعامل هذا النوع من الخوارزميات مع مشاكل الامن السيبراني من خلال توضيح طبيعة الانتقالات بين حالتي النموذج والانبعاثات التي تصدر من كل حالة مخفية ، ومن الجدير بالذكر وضع مقارنة بين النتائج بواسطة معياري المقارنة Akaike Information Criterion (AIC) و Bayesian Information Criterion (BIC) ووجد ان افضل طريقة كانت تقابل النتائج الخاصة بخوارزمية مستعمرة النحل.

معلومات البحث

تواريخ البحث:

تاريخ تقديم البحث: 7/4/2024
تاريخ قبول البحث: 26/5/2024
تاريخ رفع البحث على الموقع: 15/5/2025

الكلمات المفتاحية:

انموذج ماركوف المخفي، خوارزمية مستعمرة النمل، خوارزمية مستعمرة النحل، الامن السيبراني، التهديد المستمر المتقدم .

للمراسلة:

أحمد عبد الزهره دواي الكورجي
Ahmedalkorje0@gmail.com

doi: https://doi.org/10.55562/jrucs.v57i1.1

1. المقدمة Introduction

لدى دراسة الظواهر الطبيعية او الاقتصادية او الاتصالات او العمليات التي تحتاج خطوات مدروسة بتدبير نجد انفسنا امام حالة تتطلب دراسة متغيرات عشوائية تتغير مع الزمن، وقد وقع جزء كبير من تلك الظواهر على عاتق المختصين بالرياضيات لما لهذه الظواهر من علاقة وطيدة بنظرية الاحتمالات التي تعتبر احد اهم فروع الرياضيات التي لا تنفصل عن الحياة ولا عن متطلبات العلوم الاخرى. اصبحت الهجمات الالكترونية اكثر انتشارا وتصدرت العديد من الهجمات عناوين الاخبار على مدار العقد الماضي، واستهدفت الشركات الصناعية والمنظمات الحكومية وتسببت في خسائر مالية كبيرة وتمكنت من اعاقة تشغيل الخدمات العامة الاساسية. ومع تطور الحداثة في تنوع الحلول ارتأينا ان نتطرق الى خوارزمية مستعمرة النمل وخوارزمية مستعمرة النحل في تحسين معلمات انموذج ماركوف المخفي والتعرف عن كثب على امكانية كل طريقة واسلوبها الخاص في تحليل ونمذجة البيانات المتوفرة.

2. تحسين المعلمات بواسطة خوارزمية مستعمرة النمل [3][6][11]

المشكلة الرئيسية المتعلقة ب ACO هي تقدير معلمات النموذج من خلال التعلم من بيانات التدريب. ونظرا للأهمية العالية لمجال الذكاء الاصطناعي في وقتنا الحالي ودوره الفعال في حل مشكلات ربما تكون معقدة لذا فان خوارزمية مستعمرة النمل تعتبر من الطرق العشوائية وبديل جيد وحل اخر يرافق ما تقدمنا به من تقدير المعلمات عن طريق نظرية بيز. حيث ان تحسين مستعمرة النمل تحاكي طريقة البحث عن الغذاء لمستعمرات النمل . تتحكم مستعمرات النمل في سلوك البحث عن الغذاء دون أي ادارة مركزية وفقط عن طريق الفرمونات المترسبة على المسارات. يشار الى مسارات الفرمونات على انها وسائط اتصال معلومات والتي يشارك النمل معلوماتهم عبرها . ايضا تمثل المشاهدات هنا كالاتي $Y = y_1, y_2, \dots, y_n$ ويفترض انها تنبعث من

الحالات المخفية $X = x_1, x_2, \dots, x_n$ ويتم تحديد نموذج ماركوف المخفي من خلال المعلمات $\lambda = (\pi, A, B)$. لدينا بعض الرموز ذات الصلة مثل $Q = \{q_1, q_2, \dots, q_n\}$ وتمثل مجموعة انبعاثات وايضا لدينا $\sum = \{c_1, c_2, \dots, c_m\}$ التي تمثل مجموعة الحالات [11] .

$\pi_i, \forall i = 1, \dots, n$: يمثل احتمال البدء بالحالة $\pi = P(x_1 = q_1)$ حيث n تمثل العدد الكلي للمشاهدات

$A = [a_{ij}], i, j = 1, \dots, n$: كل a_{ij} يمثل احتمال الانتقال من الحالة q_i الى الحالة q_j

$$a_{ij} = P(x_{k+1} = q_j | x_k = q_i)$$

$B = [b_{ij}], i = 1, \dots, n, j = 1, \dots, m$: حيث b_{ij} هو احتمال ان الحالة q_i تنبعث منها c_j . الانبعاث هنا يمثل المشاهدات التي تصدر من كل حالة مخفية.

$$b_{ij} = \mathrm{P}(y_k = c_k | x_k = q_i)$$

بالنظر الى معلمات النموذج فأن احتمال انبعاث المشاهدات بواسطة الحالات المخفية يتم حسابه على النحو التالي:

$$\mathrm{P}(\mathrm{Y}, \mathrm{X}|\lambda) = \pi_{x_1} \prod_{i=1}^{n} b_{x_i, y_i} a_{x_i, x_{i+1}} \quad (1)$$

بشكل عام يتم حساب احتمالية الانبعاث والتي يشار اليها بالرمز $P(Y|\lambda)$ على النحو التالي :

$$P(Y|\lambda) = \sum_{X} \mathrm{P}(\mathrm{Y}, \mathrm{X}|\lambda)$$

حيث X هو تسلسل ممكن للحالات

يتم تحديد متوسط log-likelihood لانبعاث المشاهدات من خلال : $\mathrm{Prob}=\frac{1}{N}\sum_i^N \log P(y_i|\lambda)$

يمكن اعتبار كل مسار من مصدر الغذاء الى المستعمرة كطريق للاجتياز . يكون هناك شكل بياني[1] يولده النمل خلال حركته يحتوي على جميع المسارات الممكنة . يمشي النمل من خلال الشكل البياني وفقا للمسافة وكمية الفرمونات على كل مسار وبالتالي يمكن صياغة احتمال الاجتياز من العقدة X الى Y على النحو التالي [11][3]:

$$P_{x,y} = \begin{cases} \dfrac{(\tau_{x,y})^{\alpha} . (\eta_{x,y})^{\beta}}{\sum_{z \in N_x} (\tau_{x,z})^{\alpha} . (\eta_{x,z})^{\beta}} & y \in N_x \\ 0 & other\ wise \end{cases} \quad (2)$$

حيث N_x هي مجموعة من العقد المجاورة ل x

$\tau_{x,y}$: هي كمية الفرمون على المسار من x الى y .

$\eta_{x,y}$: هي الرغبة في الانتقال من العقدة x الى y (عادة ما تكون الرغبة هي عكس المسافة)

α, β : هما معلمتان للتحكم في تأثير المتغيرات .عندما يتحرك جميع النمل خلال الرسم البياني سيتم تحديث كميات الفرمون على النحو التالي:

$$\tau_{x,y} = (1 - \rho)(\tau_{x,y} + \sum_{\upsilon} \Delta_{x,y}^{\upsilon})$$

حيث ρ هو معامل تلاشي الفرمون

$\Delta_{x,y}^{\upsilon}$: هو كمية الفرمون التي ترسبها النملة على المسار من x الى y . تشير القيمة الكبيرة لـ ρ الى التلاشي السريع.

3. تحسين المعلمات بواسطة خوارزمية مستعمرة النحل [24]

- **انتاج مواقع اولية لمصدر الغذاء**: اذا تم اعتبار فضاء البحث هو البيئة التي توجد بها الخلية التي تحتوي على مواقع لمصدر الغذاء، تبدأ الخوارزمية بإنتاج مواقع لمصادر الغذاء بشكل عشوائي والتي تطابق الحلول في فضاء البحث. ويتم انتاج مصادر الغذاء الاولية بشكل عشوائي ضمن نطاق حدود المعايير [24]

$$X_{ij} = X_j^{min} + rand(0,1)\left(X_j^{max} - X_j^{min}\right), i = 1, \dots, S_N, j = 1, \dots, D \quad (3)$$

S_N هو عدد من المصادر الغذائية و D هو عدد قيم الامثلية او الابعاد لفضاء البحث وبالإضافة الى ذلك العدادات التي تخزن عدد التجارب من الحلول يتم اعادة تعيين قيمتها في هذه المرحلة الى 0 بعد التهيئة الابتدائية للقيم. يتعرض المجتمع من مصادر الغذاء(حلول) لتكرار دورات من عمليات البحث من النحل العامل والنحل المشاهد والنحل الكشفي ، معايير الانهاء للخوارزمية يمكن التوصل الى عدد محدد من الدورات او الوصول الى ما يطابق اقل نسبة خطأ.

- **ارسال النحل العامل الى مواقع مصدر الغذاء** : كما ذكر سابقا ان كل نحلة عاملة ترتبط مع موقع واحد لمصدر الغذاء فقط. وبالتالي فأن عدد المواقع لمصادر الغذاء هو مكافئ لعدد النحل العامل ، النحلة العاملة تحدث التعديل على الموقع

[1] لا يوجد شكل محدد يتم توضيحه هنا وانما اراد الباحث في هذا المصطلح ان يقرب الصورة للقارئ من خلال حركة النمل التي تشكل رسما بيانيا من المستعمرة الى مصدر الغذاء

لمصدر الغذاء(الحل) في ذاكرتها اعتمادا على المعلومات المحلية (المعلومات البصرية) وتجد مصدر الغذاء المجاور ومن ثم تقييم جودتها وفي هذه الخوارزمية يتم ايجاد مصدر الغذاء المجاور حسب المعادلة التالية[24] :

$$v_{ij} = X_{ij} + \emptyset_{ij}(X_{ij} - X_{Kj}) \quad (4)$$

المنطقة المجاورة لمصدر الغذاء تتمثل ب X_i ، مصدر الغذاء v_i تتعين قيمته بالتغيير على قيمة X_i . j هو عدد صحيح عشوائي في النطاق [1,D] و k هو مؤشر يتم اختياره عشوائيا ونطاقه $[1,\ldots,S_N]$ ويختلف عن i ، $\emptyset_{ij}$ وهو رقم عشوائي حقيقي في النطاق [-1,1] .

الفرق بين القيم X_{ij} و X_{Kj} يتناقص. وكما في طرائق البحث عن الحل الامثل في فضاء البحث ، طول الخطوة يتناقص تكيفيا . اذا كانت القيمة التي تنتجها هذه العملية تتجاوز حدوده المحددة سلفا، يمكن اعاده تعيين القيمة الى قيمة مقبولة . يتم تعيين القيمة اذا تجاوزت حدودها الى :

$$if\ X_i > X_i^{max}\ then\ X_i = X_i^{max}; if\ X_i < X_i^{min}\ then\ X_i = X_i^{min}$$

بعد انتاج v_{ij} ضمن الحدود، يمكن تعيين قيمة الملاءمة (اللياقة) للحل v_{ij} في مسألة الحد الادنى minimization problem [24]

$$fitness_i = \left\{\frac{1}{1+f_i}\ if\ f_i \geq 0 \middle| 1 + abs(f_i)\ if\ f_i < 0\right\} \quad (5)$$

حيث ان f_i هي دالة الكلفة costing function للحل v_{ij} . لمسألة الحد الاقصى دالة الكلفة يمكن استخدامها مباشرة بوصفها دالة لياقة . يتم تطبيق عملية اختيار الافضل greedy selection process بين X_i و v_i ، ثم يتم اختيار افضل واحد اعتمادا على قيم اللياقة التي تمثل كمية الرحيق من مصادر الغذاء X_i و v_i اذا كان مصدر في v_i هو اعلى من X_i من حيث الربحية ، والنحلة العاملة تخزن في ذاكرتها الموقع الجديد وتنسى القديم . والا يتم الاحتفاظ بالموقع السابق في الذاكرة اذا لم يتم تحسين X_i ، كما يتم زيادة العداد الذي يحسب التجارب ب1 والا تتم اعاده تعيينه الى 0.

- **حساب قيم احتمال المشاركة في الاختيار الاحتمالي**: بعد ان يكمل كل النحل العامل عمليات البحث الخاصة بهم، يقوم بمشاركة المعلومات المتعلقة بكمية الرحيق ومواقع مصادر الغذاء مع النحل المشاهد في المنطقة المخصصة للرقص. هذا النوع من التفاعل المتعدد هو ميزة خوارزمية النحل الاصطناعي حيث ان النحلة المشاهدة تقوم بتقييم معلومات الرحيق المأخوذة من كل النحل العامل وتختار موقع مصدر الغذاء بالاعتماد على الاحتمالية التي تتعلق بكمية الرحيق . هذا الاختيار الاحتمالي يعتمد على قيمة اللياقة للحلول في المجتمع . طريقة الاختيار القائم على اللياقة قد يكون باستخدام عجلة الروليت، او نظام تحديد آخر . في هذه الخوارزمية طريقة الاختيار باستخدام عجلة الروليت وكما موضح في المعادلة التالية:

$$P_i = fitness_i / \sum_{i=1}^{S_N} fitness_i \quad (6)$$

في هذه الطريقة للاختيار الاحتمالي، تزداد كمية الرحيق لمصادر الغذاء (لياقة الحلول) وبالمقابل عدد النحل المشاهد تزداد زيارتهم لتلك المصادر ايضا . هذه هي ميزة ردود الفعل الايجابية من خوارزمية النحل الاصطناعية.

- **اختيار موقع مصدر الغذاء من قبل النحل المشاهد على اساس المعلومات المقدمة من النحل العاملات [24]**: يتم توليد عدد حقيقي عشوائي ضمن نطاق [1,0] لكل مصدر. اذا كانت القيمة الاحتمالية P_i في المعادلة (6) المتعلقة بذلك المصدر هي اكبر من هذا الرقم العشوائي فالنحل المشاهد يقوم بالتعديل على هذا الموقع لمصدر الغذاء عن طريق استخدام المعادلة (4) كما هو الحال بالنسبة للنحلة العاملة. بعدها يتم تقييم المصدر بتطبيق اختيار الافضل فالنحل المشاهد اما يخزن الموقع الجديد مع نسيان القديم او يبقي الموقع القديم. اذا لم يتم تحسين الحل i ، يتم زيادة العداد الذي يحسب التجارب ب1 والا تتم اعادة تعيينه الى 0 . وتتكرر هذه العملية حتى يتم توزيع جميع النحل المشاهد على مواقع مصادر الغذاء.
- **معايير التخلي عن المصدر:** في الدورة الواحدة بعد ان تكمل كل من النحل العامل والنحل المشاهد عمليات البحث الخاصة بهم، فان الخوارزمية تدقق لمعرفة اذا كان هناك أي مصدر مستنفذ يمكن التخلي عنه. من اجل ان تقرر التخلي عن المصدر، يستخدم العداد الذي تم تحديثه خلال البحث. عداد التجارب trail اذا كانت قيمة العداد اكبر من معيار السيطرة على الخوارزمية، المعروفة باسم الحد limit ، فلأن المصدر المرتبط بهذا العداد من المفترض ان يكون مستنفذ ويتم التخلي عنه. كما هو معروف عن خوارزمية النحل الاساسية توظف معيارا واحدا فقط للسيطرة ، وهو ما يسمى الحد، ومعادلته هي :

$$limit = (S_N * MCN) \quad (7)$$

4. آلية المفاضلة

بعد ان يتم تحسين المعلمات بالخوارزميتين تلي هذه المرحلة عملية مهمة جدا وهي اختيار افضل طريقة مستخدمة حيث يتباين تمثيل البيانات ونمذجتها من طريقة الى اخرى لذا يحتم علينا ان يكون لنا قرار بخصوص أي طريقة سوف تكون افضل من حيث الكفاءة والدقة ، ويتم ذلك وفق معايير معينة حيث تم استعمال اثنين من المعايير لهذا الغرض والموضحين كالاتي :

1. **معيار معلومات اكايكي Akaike Information Criterion (AIC)**

$$AIC = 2K - Ln(L)$$

2. **معيار المعلومات البيزي Bayesian Information Criterion (BIC)**

$$BIC = KLn(T) - 2Ln(L)$$

حيث ان L هو الامكان Likelihood، T هو عدد المشاهدات ، و K هو عدد الحالات .
ان الطريقة التي يقابلها اقل قيمة من المعيارين تعتبر افضل طريقة.

5. البيانات (Data)

نظرا لأهمية الموضوع والسرية التامة التي يتعامل بها اصحاب العلاقة في اقسام الامن السيبراني في مختلف بلدان العالم ويرجع هذا الى الحساسية العالية التي ترافق المعلومات الدولية في حال تم تسريبها لما فيها من تهديد وخطورة على مفاصل الدولة السياسية والاقتصادية والاجتماعية لذا فان العاملين في هذا المجال وجب عليهم التحفظ عن اعطاء أي معلومات او خطط لهجمات سيبرانية حقيقية، وعند مراجعتنا بشكل دوري لوزارة الاتصالات العراقية وقسم الامن السيبراني على وجه الخصوص تبين ان هذا القسم قد تشكل منذ فترة لا تزيد عن سنتين وما زال في طور انشاء البنى التحتية الرقمية من خلال التعامل مع شركات خاصة وزج الموظفين في دورات تدريبية وتطويرية تتيح لهم الامكانية والجهوزية التامة لكي يتمكنوا من صد الهجمات الإلكترونية والسيطرة عليها، بالتالي فان قسم الامن السيبراني في وزارة الاتصالات بالوقت الحالي لا يمتلك قاعدة بيانات يتم الاعتماد عليها في عملنا هذا لأنه في طور الانشاء، مما دعا ذلك الى اللجوء الى وسائل اخرى غير محلية تكون فيها البيانات متاحة للجمهور والباحثين ومن يهتم بذلك. توجد انواع عديدة من الهجمات السيبرانية تتجسد من خلال رسالة نصية او برامج ضارة او غير ذلك من الادوات التي تسهل عمل المهاجم لاختراق مؤسسة معينة والاستيلاء عليها، لذا فقد تم تسليط الضوء على نوع معين من انواع الهجمات والتهديدات السيبرانية والذي يعتبر النوع الاخطر ويدعى "التهديد المستمر المتقدم" "APT" الذي تم التطرق اليه وتوضيح تفاصيله ومراحله في الفقرات السابقة .تقوم بعض المواقع الرسمية الالكترونية والشركات المعنية بنشر تفاصيل الهجوم السيبراني واهمها التي تدعى بـ "Mandiant" وهي شركة امريكية للأمن السيبراني تابعة لشركة Google حيث برزت في الصدارة في فبراير 2013 عندما اصدرت تقريرا يورط الصين مباشرة في التجسس الالكتروني. مجموعة البيانات المستخدمة في هذا العمل هي بيانات متاحة للجمهور تسمى APT-Malware حيث قامت الباحثة " Emaan Jalal Khalifa "(20) عام 2022 في كلية علوم الحاسبات جامعة ديالى بالاعتماد على هذه البيانات حيث اقترحت نظامين لتصنيف البرامج الضارة APT باستخدام التعلم الآلي المتقدم. وتتضمن 4449 عينة موزعة على عدة انواع من الهجمات الضارة التابعة الى مجموعات APT كالاتي :

$$(Equation\ Group\ ,APT_{21}\ ,APT_{28}\ ,APT_{30}\ ,APT_{1}\ ,Dark\ Hotel\ ,APT_{29}\ ,Energetic\ Bear\ ,$$
$$winnti\ Group\ ,APT_{10}\ ,APT_{19}\ ,Gorgon\ Group)$$

حسب خبراء التخصص وأصحاب الرأي والعلاقة في قسم الامن السيبراني مع وجهة نظر الباحث تم اعتماد مجموعات APT آنفة الذكر والذي يبلغ عددها 12 مجموعة كمشاهدات تنبعث من كل حالة مخفية ، في حين ان الحالات المخفية هنا نوعان فقط وهما :حالة تهديد مكتملة الخطوات وتابعة الى APT ويرمز لها بـx_1. وحالة تهديد غير مكتملة الخطوات وغير تابعة الى APT ويرمز لها بـx_2. الجدول ادناه يوضح تفاصيل حالات التهديد بصورة اكثر دقة ووضوح ليتسنى لنا فهمها وتناولها بطريقة تحدد كل مجموعة من APT وفقا لبلد المنشأ وسنة حصولها اضافة الى نوع المشاهدات المتوفرة.

جدول (1): يبين تفاصيل بيانات تهديد مجموعات APT

Country	APT Group (Y_n)	Total of threats (X)	Complete threat(x_1)	Failed threat(x_2)
China	APT_1	1007	405	602
China	APT_{10}	300	244	56
China	APT_{19}	33	32	1
China	APT_{21}	118	106	12
Russia	APT_{28}	230	214	16
Russia	APT_{29}	281	281	0
China	APT_{30}	164	164	0
North Korea	$Dark\ Hotel$	298	273	25
Russia	$Energetic\ Bear$	132	132	0
USA	$Equation\ Group$	395	395	0
Pakistan	$Gorgon\ Group$	1085	961	124
China	$winnti\ Group$	406	387	19
TOTAL		4449	3594	855

6. الجانب التطبيقي

تمثل التهديدات المستمرة المتقدمة APT وحملات الهجوم التي تنفذها جهات فاعلة مختصة وواسعة الحيلة خطرا امنيا جسيما وهناك حاجة الى ادوات مناسبة لاكتشافها، وفي هذا المجال نصف نموذجا عشوائيا لتطور APT يعتمد على نموذج ماركوف المخفي، الهدف من هذا النموذج هو التحقق مما اذا كان هناك تطور في حملات الهجوم ام لا. في هذا العمل يتم تقييم الحملات الهجومية المحتملة وتحديد ما اذا كانت حملة هجومية بالفعل بناء على فئة التهديد او الهجوم التي تحتوي عليها . يقوم المهاجم بتنفيذ عملية APT وهذا يعني ان لديهم هدفا محددا وانهم يتبعون المراحل الخاصة بالتهديد حيث يمكن للمهاجم في أي وقت ان يبدأ حركة جانبية على الشبكة الداخلية عن طريق العودة الى مرحلة الاستطلاع ولكن على جهاز مختلف. في كل مرحلة يكون المهاجم قادرا على استغلال الادوات الموجودة او نشر ادوات ضارة خاصة به، وبشكل عام فهم قادرون على اختراق أي نظام يتم استهدافه. في حالتنا هذه يتم استخدام نموذج ماركوف المخفي لنمذجة التهديدات المستمرة المتقدمة APT وهذا يعني ان الحالات المخفية تتمثل بنشاط تهديد APT وهو هل (العملية مكتملة الخطوات وتابعة الى APT) ام ان (العملية غير مكتملة الخطوات وغير تابعة الى APT) . للوقوف على عملية التقدير ودقتها، تم الاعتماد على بيانات APT كمجال تطبيقي لأنموذج ماركوف المخفي، حيث ان الخوض في هذا الجانب يعتبر مهما جدا من خلال بناء انموذج احصائي يرفد الجهات المستفيدة وخاصة وزارة الاتصالات / قسم الامن السيبراني بشيء من الافكار والسلوكيات تجاه هكذا نوع من التهديدات. اذ ان النموذج الناتج يوضح امكانية عملية الانتقال من حالة مخفية الى حالة مخفية اخرى وفق توزيع احتمالي عشوائي استنادا لما ينبعث من مشاهدات من كل حالة مخفية حيث يتم تقدير مصفوفة الاحتمالات الانتقالية المتكونة من حالتين مخفيتين اساسيتين في عملنا هذا ومعرفة فيما اذا كان من المحتمل ان تتغير أي حالة منهما الى الاخرى بقيمة احتمالية متساوية او مختلفة ومعرفة ايضا فيما اذا كانت الحالة المخفية باقية دون أي انتقال ، ويقاس ذلك من خلال تمتع الحالة المخفية بقيمة احتمالية عظمى فمن المؤكد ان احتمالية انتقالها الى الحالة الاخرى يكون ضعيفا او بقيمة احتمالية ضئيلة. وايضا مصفوفة المشاهدات B والتي سوف توضح لنا كل حالة مخفية واحتمالية وجودها ضمن أي مشاهدة.

وصف النموذج

وصف نموذج ماركوف المخفي من حيث الحالات المخفية والانبعاثات التي تصدر من كل حالة وعند تحديد النموذج علينا ان نأخذ بعين الاعتبار الاتي:

1. عدد الحالات والمشاهدات

ان الانموذج المستخدم يتكون من عدد محدد من الحالات ينبعث من كل حالة مشاهدات بتوزيع احتمالي معين، للأنموذج مجموعة من الانتقالات بين الحالات والتي تسمح بتغيير الحالة بعد انبعاث المشاهدات. فيمكننا ان نعتبر الحالات المخفية في عملنا هذا هي اثنتين فقط وهما :

1) حالة تهديد مكتملة الخطوات وتابعة الى APT ويرمز لها بـ x_1.

2) حالة تهديد غير مكتملة الخطوات وغير تابعة الى APT ويرمز لها بـx_2

حيث تكون مصفوفة الاحتمالات الانتقالية للحالات المخفية A بالأبعاد الاتية:

$$A=\begin{bmatrix} & \\ & \end{bmatrix}_{2*2}$$

اما المشاهدات التي تنبعث من كل حالة مخفية فهي مجموعات APT نفسها والتي يكون عددها 12 نوعا من المشاهدات وتوضح كالاتي:

$y_1 = APT_1$, $y_2 = APT_{10}$, $y_3 = APT_{19}$, $y_4 = APT_{21}$, $y_5 = APT_{28}$, $y_6 = APT_{29}$, $y_7 = APT_{30}$, $y_8 = Dark\ Hotel$, $y_9 = Energetic\ Bear$, $y_{10} = Equation\ Group$, $y_{11} = Gorgon\ Group$, $y_{12} = winnti\ Group$

2. الاحتمالات الابتدائية *Initial Probabilities*

يجب ان تحقق الشرط $\sum_{j=1}^{m} \pi_j = 1$ فمثلا اذا بدأت العملية العشوائية من الحالة x_1 فأن $\pi = [1\ 0\ 0\ \dots]$ ، اما اذا بدأت العملية العشوائية من الحالة x_2 فأن $\pi = [0\ 1\ 0\ 0\ \dots]$

هنا نفترض ان العملية العشوائية تبدأ من أي حالة من الحالات المخفية وبالتالي يكون توزيع الاحتمالات بالشكل الاتي :

$$\pi = [\frac{1}{2}\ \ \frac{1}{2}\ ]$$

اما فيما يخص المصفوفة الاحتمالية B التي تكون مصفوفة رابطة بين الحالات المخفية والمشاهدات التي تنبعث من كل حالة تكون بالأبعاد الاتية :

$$B=\begin{bmatrix} & \\ & \end{bmatrix}_{2*12}$$

ليست هناك طريقة مباشرة لاختيار الاحتمالات الابتدائية وهذا امر مهم يتبادر الى الاذهان ، ولعمل ذلك تم اختيار القيم بشكل عشوائي او تخميني بشرط تحقيق مبدأ التصادفية.

نتائج تطبيق خوارزمية النمل

- **اولا: مصفوفة الاحتمالات الانتقالية**

$$A = \begin{bmatrix} 0.99312 & 0.00688 \\ 0.09981 & 0.90019 \end{bmatrix}$$

- **ثانيا: مصفوفة الانبعاثات**

$$B = \begin{bmatrix} 0.02204 & 0.00241 & 0.00196 & 0.30628 & 0.17299 & 0.00434 & 0.01178 & 0.14735 & 0.00117 & 0.00112 & 0.00768 & 0.32087 \\ 0.01391 & 0.16422 & 0.00287 & 0.19738 & 0.18052 & 0.05159 & 0.20823 & 0.04637 & 0.00398 & 0.02551 & 0.06646 & 0.03895 \end{bmatrix}$$

- **ثالثا: متجه الاحتمالات الابتدائية**

$$\pi = [0.60822 \quad 0.39178]$$

نتائج تطبيق خوارزمية النحل

- **اولا: مصفوفة الاحتمالات الانتقالية**

$$A = \begin{bmatrix} 0.98702 & 0.01298 \\ 0.14659 & 0.85341 \end{bmatrix}$$

- **ثانيا: مصفوفة الانبعاثات**

$$B = \begin{bmatrix} 0.03402 & 0.00088 & 0.00000 & 0.13792 & 0.00014 & 0.10335 & 0.03020 & 0.13869 & 0.00335 & 0.13645 & 0.01011 & 0.40489 \\ 0.03604 & 0.16221 & 0.05784 & 0.15860 & 0.21328 & 0.00028 & 0.20582 & 0.03905 & 0.00994 & 0.00931 & 0.06561 & 0.04202 \end{bmatrix}$$

- **ثالثا: متجه الاحتمالات الابتدائية**

$$\pi = [0.62908 \quad 0.37092]$$

7. المقارنة

من خلال معياري المفاضلة الموضحين في الجدول ادناه تبين لنا انه فيما يخص معيار AIC فأن اقل قيمة كانت تقابل خوارزمية مستعمرة النحل وهي 176.6338، اما معيار BIC ايضا نلاحظ ان ادنى قيمة كانت ترافق خوارزمية مستعمرة النحل وهي 181.4829 وهذا يوضح ومن خلال فلسفة المعيارين ان افضل طريقة لتحسين معلمات نموذج ماركوف المخفي هي خوارزمية مستعمرة النحل .

جدول(2): يوضح معياري المقارنة على تطبيق البيانات

	ACO	BCO
AIC	177.2168	**176.6338**
BIC	182.0659	**181.4829**

8. الاستنتاجات Conclusions

في معادلة رقم(1) لدينا دالة امكان تتضمن المعلمات والتي تستند على الحالات المخفية والمشاهدات التي تنبعث من كل حالة، تم تطبيق خوارزمية مستعمرات النمل وخوارزمية مستعمرات النحل في محاولة لتحسين معلمات انموذج ماركوف المخفي والذي يتمثل بتعظيم دالة الامكان وظهرت لنا النتائج كما هي موضحة آنفا وبالمعلمات المذكورة من حيث طبيعة الانتقالات والانبعاثات. تظهر لنا نتائج خوارزمية مستعمرة النمل ان امكانية ان تتحول حالة التهديد غير المكتملة الى حالة تهديد تامة ومكتملة باحتمالية (0.09981) ، بينما امكانية ان تكون حالة التهديد مكتملة وتنتقل الى وضع متراجع وغير مكتمل فتكون بالاحتمال التالي (0.00688) ، وايضا نتائج خوارزمية مستعمرة النحل نجد ان تحول حالة التهديد غير المكتملة الى حالة مكتملة يصل بنسبة احتمالية 0.14659 اما التهديد المكتمل فيكون احتمال ان يتراجع الى غير مكتمل يصل بنسبة 0.01298 في حين ان بقاء كل حالة تهديد في نفس وضعها وظروفها ومكانها سوف يحتل اكبر قيمة احتمالية في كلتا الخوارزميتين. ان الجزء الاكثر اهمية في هذه النتائج هي مصفوفات الانبعاثات B حيث نلاحظ من خلال نتائجها ان الحالة الاولى (التهديد المكتمل) في خوارزمية مستعمرة النحل من الممكن ان تحقق غايتها والاستيلاء على أي مؤسسة وسرقة او تعطيل محتواها دون المرور بكل المشاهدات والمميزات استنادا الى بعض الاحتمالات التي تكون صفرية او قليلة اكثر من باقي المشاهدات مما يوضح ان مميزات التهديد المكتمل له القدرة وفق امكانياته المتطورة ان يختصر بعض الخطوات او يستغني عنها عندما يراد حياكة تهديد ضد مؤسسة معينة او جهة امنية والسيطرة على بياناتها.

9. التوصيات Recommendation

من الناحية العلمية يوصي البحث بتطبيق باقي خوارزميات الذكاء الاصطناعي وبالخصوص باقي خوارزميات ذكاء الأسرب ومعرفة ما اذا كانت النتائج تمثل الجانب التطبيقي افضل تمثيل وبصورة مناسبة وايضا التركيز على معرفة باقي انواع تهديدات الامن السيبراني واللعب على متغيرات جديدة ومهمة قد ترفد البلد بنتائج مفيدة، وهذا مجال مهم للباحثين ان يتوصلوا الى قيمة علمية تساهم في بناء اسس رصينة .اما من الناحية الفنية فمن اهم التوصيات التي يحتاج لها اصحاب العلاقة في وزارة الاتصالات

وبالخصوص قسم الامن السيبراني هو بناء جدار حماية الكتروني رصين يحد من هكذا نوع من الخطورات . بالإضافة الى جدار الحماية الرصين يجب ان تكون هناك برامج خاصة لمكافحة الفيروسات والتي وظيفتها التنبيه بالإصابة بالفيروسات والبرامج الضارة، كما سيقدم العديد منها خدمات اضافية مثل فحص رسائل البريد الالكتروني للتأكد من خلوها من المرفقات الضارة.

المصادر

[1] ازهري، نور مصطفى (2017) " استخدام طوريات ماركوف المخفية في التعرف على الصور والرموز " رسالة ماجستير ، كلية العلوم قسم الرياضيات ، جامعة تشرين، الجمهورية العربية السورية.

[2] الخياط، باسل يونس ذنون ، فاطمة محمد حسن (2011) " استخدام ثلاثة طرائق احصائية للتعرف على عدد من التغيرات في سلسلة المادة الوراثية " المجلة العراقية للعلوم الاحصائية ، ص(1-20).

[3] الدباغ، نجلاء بديع , باشي ، محمود صبحي (2012) " استخدام خوارزمية النمل في كشف وتصنيف التطفل في شبكات الحاسوب " مجلة العلوم والتربية ، العدد (2) ، ص (146-168).

[4] ديوان، اسماء حسين سمير (2016) " استخدام سلاسل ماركوف المخفية في تحليل البيانات الحيوية" اطروحة دكتوراه، كلية الادارة والاقتصاد الجامعة المستنصرية .

[5] رغد وشهباء (2014) " اختيار حالات الاختبار وتعيين اسبقيتها باستخدام مستعمرة النحل " مجلة الرافدين لعلوم الحاسوب والرياضيات ، المجلد(11) ، العدد (1).

[6] سليمان، اسماء صلاح الدين (2019) " استخدام خوارزمية مستعمرة النمل لإيجاد التخصيص الامثل " مجلة جامعة الانبار للعلوم الاقتصادية والادارية ، المجلد (1) ، العدد (25).

[7] العبيدي، عبدالغفور(2018) " استخدام سلاسل ماركوف في المجالات الطبية " المجلة العراقية للعلوم الاحصائية، العدد(25).

[8] قاسم، عمر صابر (2014) " تهجين انموذج ماركوف المخفي باستخدام شبكة ايلمان العصبية الاصطناعية مع التطبيق " مجلة الرافدين لعلوم الحاسوب والرياضيات ، العدد(1) ، المجلد(11).

[9] مدلول ، امير ساجت (2022) " تأثير التحسين المستمر في جودة الامن السيبراني بتوسيط التوجه نحو الزبون " رسالة ماجستير، جامعة القادسية .

[10] Aithal,Gaikwad and sahve(2015)," speech enhancement using PCA for speech and emotion recognition" Global journal of engineering , vol.4(3) , P(6-12).

[11] Akram emdadi and others(2018)," A noval algorithm for parameter estimation of hidden markov model inspired by ant colony optimization " Heliyon 5(2019) , doi: 10.1016 / j.heliyon .

[12] Anders krogh and Larsson (2001)," Predicting trans membrane protein topology with a hidden markov model: application to complete genomes " journal of molecular Biology , vol.305, ISSUE 3, P(567-580).

[13] Andre lnge(2013)," Hidden markov models theory and simulation" thesis , mathematical statistics, Stockholm university.

[14] Anton Tenyakov(2014), " Estimation of hidden markov models and their Applications in finance " Electronic Thesis and dissertation repository.

[15] Bagos and Hamodrakas(2006)," Algorithms for incorporating prior topological information in HMMs: application to trans membrane proteins " Bio med central Bioinformatics.

[16] Birney(2001)," Hidden markov models in biological sequence analysis " IBM journal of research and development , vol(45) , Issue(3.4) , P(449-454).

[17] Brogi and Bernavdino(2019)," hidden markov models for advanced persistent threats" Article in international journal of security and networks, vol.(14),No.(4).

[18] Buyers Guide (2022)," Advanced persistent threat (APT)" GSA , VERSION 2.0.

[19] Chenquan and others (2023)," Advanced persistent threats and Their defense methods in industrial internet of things:Asurvey " Mathematics 2023 , 11,3115 , https ;// doi.org/ 10.3390 / math 11143115.|

[20] Christophe chesneau ,Salima elkolei(2022)," Parametric estimation of hidden markov models by least squares type estimation and deconvolution " statistical papers , springer verlag , hal-01598922v6

[21] Eman and Dhahair (2022)," Concept and difficulties of advanced persistent threats (APT): Survey " Int.j.nonlinear Anal , APPL. 13(2022)1 , 4037-4052 . ISSN : 2008-6822.

[22] Ghafir(2019)," hidden markov models Alert correlations for the prediction of Advanced persistent threats " Article in IEEE Access , digital object identifier.

[23] Guillaume Brogi(2018)," Real-time detection of Advanced persistent threats using information flow tracking and hidden markov models " Thesis Doctora.

[24] Jagdish and others (2013)," Artifical bee colony algorithm: a survey " Int. journal intelligence paradigms , vol. (5).

[25] Lei and chen (2017)," Detect Advanced persistent threat in graph –level using competitive Auto encoder" P(35-41),communications and information technology.

المواقع الالكترونية

[26] Threat Group Cards: A Threat Actor Encyclopedia, available at: https//apt.etda.or.th/cgi-bin/show card.

[27] Mandiant, American cyber security firm, available at: Threat Intelligence Solutions | Cyber Security Services & Training

Journal of AL-Rafidain University College for Sciences (2025); Issue 57; pp. 1-9

AL- Rafidain University College

PISSN: (1681-6870); EISSN: (2790-2293)

Journal of AL-Rafidain University College for Sciences

Available online at: https://www.jrucs.iq

JRUCS

Journal of AL-Rafidain University College for Sciences

Intelligent Methods for Estimating Hidden Markov Model Parameters

Ahmed A. Douai Al-Korgi	Prof. Dr. Hamid S. Nour Al-Shammari
Ahmedalkorje0@gmail.com	Hamed_Saad@albayan.edu.iq
Department of Statistics, College of Administration and Economics, Al-Mustansiriya University, Baghdad, Iraq	College of Business Administration, Al Bayan University, Baghdad, Iraq

Article Information

Article History:
Received: April, 7, 2024
Accepted: May, 26, 2024
Available Online: May, 15, 2025

Keywords:
Hidden Markov model, ant colony algorithm, bee colony algorithm, cyber security, advanced persistent threat.

Correspondence:
Ahmed A. Douai Al-Korgi
Ahmedalkorje0@gmail.com

Abstract

Hidden Markov models (HMMs) are stochastic models that were initially applied as statistical models for speech and handwriting recognition because of their great ability to adapt to the problem and their skill in dealing with sequential signals. With the development of techniques, tools, and methods for estimating the parameters of the hidden Markov model, attention may turn to smart methods due to their importance and wide use among researchers. Two algorithms were taken, namely the Ant Colony Optimization (ACO) algorithm and the Bees Colony Optimization (BCO) algorithm. They were applied to an important field at present, which is cyber security, where we addressed one of the threats that pose a danger, which is the Advanced Persistent Threat (APT). The results showed the flexibility in dealing with this type of algorithm for cyber security problems by clarifying the nature of the transitions between the two model states and the emissions that come from each hidden state. It is worth noting that a comparison of the results was made using the two comparison standards, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), and we found that the best method was the results of the bee colony algorithm.

doi: https://doi.org/10.55562/jrucs.v57i1.1