

استخدام الخوارزمية الجينية لتقدير نقاط التغيير في إحدى نماذج النظم الحركية غير المستقرة مع التطبيق

د. ظافر رمضان مطر البدراني
thafer63@yahoo.com

د. نجلاء سعد ابراهيم الشرايبي
alialialiab@yahoo.com

الملخص :

تم في هذا البحث استخدام إحدى أساليب التشخيص للأنظمة الحركية غير المستقرة، وهو أسلوب تقطيع البيانات إذ تبدأ معالجة مشكلة تشخيص النظم غير المستقرة من التشخيص خارج الخط الذي يعد الحجر الأساس في الوصول إلى النموذج الذي يعد الأكثر ملاءمة في وصف النظام ، من بعدها يتم إجراء عملية التقطيع للبيانات ، إذ يتم تقطيع سلسلتي المدخلات والمخرجات إلى فترات زمنية بإذ يمكن الحصول على بيانات مستقرة ضمن كل فترة تقريبا. الهدف الأساسي لعملية تقطيع البيانات هو تقليل دالة الخسارة ، وكذلك الحصول على نموذج يمثل فترة البيانات المعنية أفضل ما يمكن. تم توظيف الخوارزمية الوراثية لإيجاد عدد ومواقع نقاط التغيير وإيجاد الأنموذج الملائم لكل قسم باستخدام التطبيق على بيانات مولدة، وبينت النتائج التي حصل عليها البحث بالتطبيق أنها طريقة جيدة وكفؤة في تقدير مواقع وعدد نقاط التغيير.

الكلمات المفتاحية: الخوارزمية الجينية ، نماذج النظم الحركية

Using Genetic Algorithm for Estimating Change Point in One Non-Stationary Dynamic Systems Model with Application

Abstract:

One of the diagnostic methods for non-stationary dynamic systems was used in this research , namely data segmentation. The identification problem for non-stationary systems starts from the outline diagnostic, which considered as the cornerstone of reaching a model that is more appropriate in describing the system . Consequently the data segmentation operation is performed, were the input and output series are segmented into time intervals so that stable data can be obtained within each interval. The main objective of the data segmentation process is to minimize the loss function and obtain a model that represents the data interval in question as best as possible. We used the genetic algorithm to find the number, location change points, and model fits for each part by applying a generator. Results showed that it is a good and efficient way to estimate the location and number of points of change.

*مدرس / قسم الإحصاء والمعلوماتية / كلية علوم الحاسوب والرياضيات / جامعة الموصل

** استاذ / قسم الإحصاء والمعلوماتية / كلية علوم الحاسوب والرياضيات / جامعة الموصل

تاريخ القبول 2018/9/26

تاريخ استلام البحث 2018/9/10

(1) المقدمة:- Introduction

يظهر النظام أحياناً تغييرات مفاجئة في سلوكه عبر الزمن عندما تكون البيانات متجمعة (أي جمعت مسبقاً) ، وللكشف عن هذه التغييرات وتحديد مواقعها فقد استخدم إجراء التقطيع (segmentation procedure) ، و طبقت هذه المسألة في مجالات عدة منها دراسة الجهاز العصبي وطب القلب وتمييز الصوت،...الخ . والسبب الرئيس لاستخدام مسألة التقطيع في المجالات المذكورة في أعلاه هو أنَّ سلوك النظام الحركي يكون غير مستقر، فيتم تقطيع هذا السلوك إلى فترات زمنية أطلق عليها تسميات عدة منها التقسيمات (Partitions) أو القطاعات (Blocks) أو القطع (Segments) بحيث يصبح هذا السلوك ضمن كل قطاع مستقراً تقريباً. عموماً فإن مسألة التقطيع تهدف للإجابة عن التساؤلات الآتية: (Badagian, 2013)

أ- هل توجد تغييرات ؟

ب- متى حدثت هذه التغييرات؟

ت- إذا حدث أكثر من تغيير واحد كيف يمكن تحديد أماكن هذه التغييرات؟

للإجابة عن الأسئلة في أعلاه استخدمت معايير المعلومات الإحصائية لاكتشاف وتقدير وتحديد مواقع نقاط التغيير، إذ توجد عدة نظريات لحل هذه المسألة جميعها مستندة على الطرائق المعلمية وغير المعلمية ، وذلك عن طريق مقارنة الإحصائيات المحسوبة تحت فرضية انه لا توجد تغييرات في النظام الحركي. والفكرة الأساسية لعملية تقطيع النظام هو تقليل دالة الخسارة (Lost Function) وبناء نموذج رياضي يلائم تلك البيانات. (Davis et.al , 2006)

لتكن (u_t, y_t) مدخلات ومخرجات النظام الحركي تحتوي على (m) من نقاط التغيير في المواقع $(k_1^*, k_2^*, \dots, k_m^*)$ إذ أن $(1 \leq k_1^* \leq k_2^* \leq \dots \leq k_m^* \leq T)$ فإن فرضية الاختبار هي:

$$H_0 : \theta_1 = \theta_2 = \dots = \theta_m = \theta$$

$$H_1 : \theta_1 = \theta_2 = \dots = \theta_{k_1^*} \neq \theta_{k_1^*+1} = \theta_{k_1^*+2} = \dots = \theta_{k_2^*} \neq \dots \neq \theta_{k_{m-1}^*+1} = \theta_{k_{m-1}^*+2} = \dots = \theta_{k_m^*} \neq \theta_{k_m^*+1} = \dots = \theta_T$$

يمكن أن تكون قيم $(\theta_i ; i=1,2,\dots,m)$ معلومة او غير معلومة ، إذ m تمثل عدد نقاط التغيير، ثم إن الهدف الأساسي هو اكتشاف وتحديد المواقع $(k_1^*, k_2^*, \dots, k_m^*)$ وأيضاً تقدير قيم (θ_i) عندما تكون غير معلومة.

(2) بعض المفاهيم الأساسية:

1- الخوارزمية الجينية Genetic Algorithm: تعرف الخوارزمية الجينية بأنها طريقة للانتقال من السكان الأوائل (الكروموسومات Chromosomes) إلى سكان جدد باستخدام نوع من الاختيار الطبيعي مع المشغلين الملهمين بعلم الوراثة وهما العبور (crossover) والطفرة (Mutation). وكل كروموسوم يحتوي على جينات (Genes) ، وكل جين يُكوّن حالة معينة.

ويختار مشغل الاختيار الكروموسومات من السكان التي سيُسمح لها بإعادة الإنتاج ، إذ تختار الكروموسومات الأكثر لياقة لكي تنتج نسلًا أكثر لياقة. (Davis et.al , 2006)

2- المجتمع population: وهو مجموعة من الأفراد المولدة بشكل عشوائي ، وكل فرد يتم تقييمه من دالة التقييم المجهزة من المبرمج ، ويعطى لكل فرد نتيجة تستند على أدائه تجاه المهمة المعطاة. إذن يتم اختيار الفردين بالاستناد على قيمة لياقتهم ، أي الذين لهم أعلى لياقة ولهم الفرصة الأعلى للاختيار لتوليد واحد أو أكثر من النسل. (Hamed et. al,2011)

3- الكروموسوم Chromosomes : هو سلسلة من الجينات التي تمثل الحل (الحلول). (Rojas,1996)

4- الجين Gene: هو الجزء الوحيد الذي يتم ترميزه من فضاء الحلول، بمعنى آخر هو القطعة الوحيدة أو القطع الصغيرة من القطع المجاورة التي ترمز الى العنصر المرشح للحلول. (Cheng & Xiaoyong, 2011)

5- دالة اللياقة Fitness Function: هي الدالة التي تحدد الحل المثالي (الكروموسوم) ، ذلك الحل المعين الذي يصنف ضد كل الحلول الأخرى. التي تقيس نوعية الحل الأمثل ودائماً تتابع المسألة المشكلة. (Amiya et. al,2012)

(3) القواعد الأساسية لتطبيق الخوارزمية الوراثية:-

Rules Base Application Genetic Algorithm

إذا أردنا تطبيق خوارزمية الجينات الوراثية هنالك اربع نقاط مهمة: (Rojas,1996)
أولاً: الترميز Encoding

ترميز الكروموسوم هو أول خطوة في حل المسألة ، ويعتمد كلياً على المسألة لتمثيل الحلول على شكل خيوط تحمل المعلومات الضرورية، كل جين في الكروموسوم يسيطر على خاصية معينة من الفرد بالطريقة نفسها، فكل قطعة في الخيط تمثل خاصية للحل. ويكون ترميز الفرد بإحدى الطرائق الآتية: (Cheng & Xiaoyong, 2011)

1- الترميز الثنائي (Binary Encoding) ويعد من الطرائق الشائعة لترميز الكروموسومات وهي سلسلة من (0,1) ، إذ الحل عبارة عن مجموعة من الأصفار والوحدات التي تعبر عن حل للمشكلة.

2- ترميز التباديل بأعداد حقيقية (Permutation Encoding) ، وفيه يتم التمثيل بأرقام صحيحة أو حروف وهو سهل جداً ويستخدمه كثير من المبرمجين.

3- الترميز بسلسلة من القواعد أي (R1,R2,...).

4- الترميز من خلال برمجة العناصر أي (برمجة كيفية الترميز).

ثانياً: الاختيار Selection

هو اختيار الأفراد الأكثر منظورية الذين سيشترون في تكوين جيل السكان القادم. وتتم عملية اختيار الإباء الذين لديهم أفضل لياقة ليشتروا بتكوين الجيل التالي من خلال استخدام إحدى الطرائق وهي: (Hamburg, 2005)

1- اختيار عجلة روليت Roulette wheel Selection:

وهي دائرة مقسمة إلى (N) من القطاعات، إذ أن (N) هي عدد الأفراد في السكان ويتم تقوس كل قطاع نسبة إلى الاحتمالية المقابلة للفرد، وبعد كل دوران للعجلة الفرد المقابل لسهم الاختيار سيتم اختياره. ويمكن القول بأن في عجلة روليت يتم اختيار الحلول حسب جودتها ، فالحل ذو القيمة الأفضل نسبة اختياره في الجيل الذي ستم عليه عملية العبور الطفرة كبيرة.

2- اختيار الرتب Rank Selection:

ويتم خزن الأفراد طبقاً لقيمة دالة اللياقة ، والتي تخصص كرتبة إذ يخصص لأفضل فرد الرتبة (1) ، والفرد الأفضل الثاني بالرتبة (2) وهكذا . وتحسب احتمالية اختيار كل فرد طبقاً للدالة غير الخطية الآتية:

$$p = \beta(1 - \beta)^{(rank-1)} \quad \dots(1)$$

إذ أن: β تمثل المعلمة ، ويعرفها المستخدم.

3- اختيار النخبة Elitist Selection :

وهي طريقة لاختيار الحلول الأفضل من المجتمع (population) ، وإلغاء باقي الحلول منه ، أي اختيار الفرد الذي يمتلك أكبر قيمة لدالة اللياقة اذا كانت المشكلة هي تحقيق الحد الأقصى (maximization) ، وبالعكس اختيار الفرد الذي يمتلك اقل قيمة لدالة اللياقة اذا كانت المشكلة هي تحقيق الحد الأدنى (minimization) . وتعد هذه الطريقة الأسهل على الإطلاق. (Davis et.al , 2006)

ثالثاً: العبور Crossover

وهي العملية المسؤولة عن تبادل بعض أجزاء من الحلول بين أبوين (حليين) لينتج حلولاً جديدة ترث من هذين الحليين الأوليين. الهدف من عملية العبور هو السماح للطفل

تسلم افضل الأجزاء المختلفة من أبائه. وهنالك عدة طرائق لعملية العبور منها:
(Davis et.al , 2006)

1- العبور بنقطة واحدة Single point crossover:

و يتم فيها اختيار نقطة بشكل عشوائي من الكروموسومات الفردية ، ويتم قطع الحلين عندها ويتم تبادل المادة الوراثية عند هذه النقطة.

2- العبور بنقطتين Two point crossover:

و يتم فيها اختيار نقطتين بشكل عشوائي من الكروموسومات الفردية ويتم قطع الحلين عندهما ويتبادل المادة الوراثية بينهما.

3- العبور المنتظم Uniform crossover :

وفيهما يتم تخصيص الصور (Heads) للأب الأول والكتابة (Tails) للأب الثاني ، ويتم رمي قطعة النقود وجينات الطفل الأول تؤخذ جينات الأب الذي سيظهر عند رمي قطعة النقود ، والطفل الثاني يأخذ جينات الأب الآخر على العكس تماماً.

رابعاً: الطفرة Mutation

كروموسوم الطفل الواحد ينتج من كروموسوم أب واحد، جوهرياً الطفل سيكون له جينات هي جينات والده نفسها باستثناء عدد صغير من الجينات التي سيتم تعديلها عشوائياً. ويمكن تعريفها بأنها السلسلة التي تتغير بتعمد لكي يبقى التنوع في مجموعة السكان.
(Cheng & Xiaoyong, 2011)

(4) مبدأ أقل وصف طولي: principle Minimum Description Length(MDL)

يُعد مبدأ MDL واحداً من طرائق الاستدلال الإحصائي ، الذي يزودنا بحل عام لمسألة اختيار النموذج، يستند مبدأ MDL على البصيرة التالية ، وهي الانتظام الموجود في البيانات يمكن استخدامه لكبس البيانات والانتظام الأكثر هو كبس أكثر ما يمكن من البيانات، المنشيء الرئيس لهذا المبدأ MDL هو العالم Rissanen عام 1989. يمثل الرمز M أنموذج ARX لكامل البيانات (قبل إجراء عملية التقطيع) ، والرمز γ يمثل أنموذج ARX المقطع إذ $\gamma \in M$. إحدى الطرائق الكلاسيكية لخزن المشاهدات (البيانات) هي تقسيم البيانات إلى مكونين ، وهما ملائمة الأنموذج $\hat{\gamma}$ زائداً الجزء غير المفسر من $\hat{\gamma}$ ، ويقصد بالجزء الثاني البواقي التي يرمز لها $e = y - \hat{y}$ إذ إن $\hat{y}$ تمثل المتجه الملائم لـ y . فإذا كانت $CL(\bullet)$ تمثل الترميز الطولي للشيء فإن :

$$CL(y) = CL(\hat{\gamma}) + CL(\hat{\gamma}|y) \quad \dots(2)$$

إذ إن $CL(\hat{\gamma})$: تمثل الترميز الطولي للأنموذج الملائم $\hat{\gamma}$.

$\hat{\gamma}$: يمثل المتجه γ الذي يحتوي على المعلمات $(a_1, a_2, \dots, a_{m+1}, b_1, b_2, \dots, b_{m+1})$ ، التي تم تقديرها بطريقة المربعات الصغرى $(\hat{a}_1, \hat{a}_2, \dots, \hat{a}_{m+1}, \hat{b}_1, \hat{b}_2, \dots, \hat{b}_{m+1})$ ، وان كلاً من a_j, b_j يمثلان رتب الأنموذج الخاصة بالقطاع j إذ $j=1,2,\dots,m+1$ ، علماً بأن m معروفة مسبقاً بأنها عدد نقاط التغيير.

$CL(\hat{\gamma})$: تمثل التشفير الطولي للبواقي (e) (المشروط بالأنموذج الملائم $\hat{\gamma}$).

كل من $CL(\hat{\gamma})$ و $CL(\hat{\gamma})$ ممثلان بالمعادلتين الآتيتين، انظر (الشرابي، 2017)

$$CL(\hat{\gamma}) = \log_2(m) + (m+1)\log_2(n) + \sum_{j=1}^{m+1} \left[\log_2(a_j) + \log_2(b_j) + \frac{a_j + b_j + 2}{2} \log_2(n_j) \right] \quad \dots(3)$$

$$CL(\hat{\gamma}) = \sum_{j=1}^{m+1} \left[\frac{n_j}{2} \log_2(2\Pi) - \frac{1}{2} \log_2|\hat{V}_j| + \frac{1}{2} y_j^T \hat{V}_j y_j \right] \quad \dots(4)$$

وبتعويض قيمة $CL(\hat{\gamma})$ و $CL(\hat{\gamma})$ بالمعادلة (1) نحصل على

$$CL(y) = \log_2(m) + (m+1)\log_2(n) + \sum_{j=1}^{m+1} \left[\log_2(a_j) + \log_2(b_j) + \frac{a_j + b_j + 1}{2} \log_2(n_j) \right] + \sum_{j=1}^{m+1} \left[\frac{n_j}{2} \log_2(2\Pi) - \frac{1}{2} \log_2|\hat{V}_j| + \frac{1}{2} y_j^T \hat{V}_j y_j \right] \quad \dots(5)$$

وبأستخدام التقريب القياسي للامكان ، أي بمعنى تقريب المقدار $(\hat{\sigma}_j^2) \approx -2 \log_2(\text{likelihood}) - n_j \log_2(\hat{\sigma}_j^2)$ إذ أن $(\hat{\sigma}_j^2)$ هي القيمة المقدرة لـ (σ_j^2) الخاصة بالقطاع (j) فان مبدأ MDL يعرف بالمعادلة الآتية:

$$MDL(m, k_1, k_2, \dots, k_m, a_1, a_2, \dots, a_{m+1}, b_1, b_2, \dots, b_{m+1}) = \log_2(m) + (m+1)\log_2(n) + \sum_{j=1}^{m+1} \left[\log_2(a_j) + \log_2(b_j) + \frac{a_j + b_j + 1}{2} \log_2(n_j) \right] + \sum_{j=1}^{m+1} \left[\frac{n_j}{2} \log_2(2\Pi \hat{\sigma}_j^2) \right] \quad \dots(6)$$

(5) تفاصيل تطبيق الخوارزمية الجينية لمسألة التقطيع:-

Specification Implementation GA to the segmentation problem

تم توظيف الخوارزمية الجينية في ايجاد عدد ومواقع نقاط التغيير، ورتب النماذج ضمن القطاعات في النظم الحركية الخطية ، وذلك من خلال استناد الخوارزمية الجينية على مبدأ MDL التي ستكون من خلال الخطوات الآتية:

الخطوة الاولى: الترميز إذ سيتم ترميز المعلومات $(m, k_1, \dots, k_m, a_1, \dots, a_{m+1}, b_1, \dots, b_{m+1})$ في متجه يدعى بالكروموسوم ، إذ تدعى المكونات الموجودة داخل الكروموسوم بالجينات، إذ إن m تمثل عدد نقاط التغيير، و $(k_1, \dots, k_m)$ تمثل مواقع نقاط التغيير، و $(a_1, \dots, a_{m+1}, b_1, \dots, b_{m+1})$ تمثل رتب

أنموذج ARX الخاصة بكل قطاع تفاصيل أكثر عن عملية الترميز المعلومات المذكورة سابقاً انظر (الشرابي ، 2017).

الخطوة الثانية: تكوين المجتمع الابتدائي الذي يحتوي على N من الكروموسومات المولدة بشكل عشوائي ، ويتم تقييم كل كروموسوم من خلال إيجاد قيمة MDL له ، ثم يختار كروموسوم الأب بشكل عشوائي من هذه المجموعة من المجتمع على أساس الكروموسوم ، الذي يمتلك أقل قيمة لمبدأ MDL ، ويتم تكوين النسل بتطبيق عملية العبور أو التغير، تفاصيل أكثر عن عملية تكوين النسل انظر (الشرابي ، 2017).

الخطوة الثالثة: تعاد العملية لتوليد أجيال جديدة إذ ستستمر الخوارزمية الوراثية حتى يقترب المعيار من القيمة المثلى ، ولكي يتم الوصول إلى الحل الأمثل تم تطبيق أنموذج الجزيرة (Island Model) انظر (الشرابي ، 2017) . كذلك تم افتراض قيود كحد أعلى وهي (a_0, b_0) للرتب (a_j, b_j) الخاصتين بأنموذج ARX ضمن كل قطعة. ولتقدير الأنموذج بشكل جيد جرى وضع شرط هو أن كل قطعة يجب أن تحتوي على عدد كافٍ من المشاهدات لكي يتم تقدير الأنموذج لها بشكل صحيح ، وسمي هذا الشرط بالمدى الأدنى (Minimum Span) الذي يرمز له $sp(*)$ ، ويضعه المستخدم وفقاً للمسألة المعطاة (Richard et. al., 2008). ويعتمد عدد المشاهدات داخل القطعة (j) على الرتب (a_j, b_j) إذ يكونان بشكل دالة متزايدة $(a_j + b_j)$ التي يتم تمثيلها بـ $sp(a_j + b_j)$.

الخطوة الرابعة: تكرار الخوارزمية لعدد من التكرارات numbiter الذي يتم على أساسه تحديد عدد نقاط التغير ضمن النظام ، وذلك من خلال حساب النسب المئوية لعدد القطاعات.

(6) تجارب المحاكاة: Simulation Experiments

لقد تم تطبيق الخوارزمية الجينية لإيجاد عدد ومواقع نقاط التغير ، والرتب الخاصة بنماذج كل قطاع على بيانات مولدة وفق النماذج الآتية:

أولاً: التطبيق على بيانات مولدة وفق الأنموذج L1 بعدد بيانات (500) مشاهدة . وبنقطة تغيير واحدة. فقد تم توليد أنموذج ARX(1,2,1) لبيانات القطاع (القسم) الأول وأنموذج ARX(1,2,1) لبيانات القطاع الثاني. والصيغة العامة للأنموذج الذي تم توليد البيانات على أساسه موضح بالمعادلة الآتية:

$$L1: y_t = \begin{cases} 0.6y_{t-1} + 0.3u_{t-1} - 0.2u_{t-2} + e_t & \text{if } 1 \leq t \leq 300 \\ 0.96y_{t-1} + 0.4u_{t-1} - 0.5u_{t-2} + e_t & \text{if } 301 \leq t \leq 500 \end{cases}$$

إن المدخلات u_t هي عبارة عن إشارة عشوائية (Random Signal) مولدة من التوزيع الطبيعي القياسي $N(0,1)$. وهناك أنواع مختلفة من إشارات الإدخال اعتمدنا في دراستنا على نوع

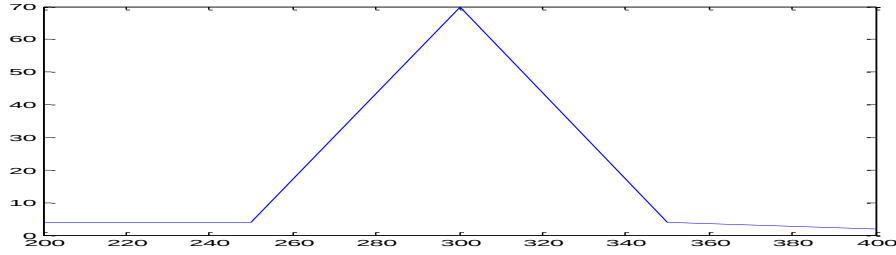
المسماة rgs (Random Gaussian Signal) للتجهيز بالمدخلات. أما التشويش e_t فقد تم توليده من التوزيع الطبيعي القياسي $N(0,1)$ ، واستخدمت الدالة randn للتوليد من التوزيع الطبيعي القياسي $N(0,1)$. لتنفيذ الخوارزمية الوراثية لإيجاد عدد ومواقع نقاط التغيير وكذلك النماذج الخاصة بكل قطاع بعد تقطيع النظام عند مواقع نقاط التغيير يجب تحديد بعض الاحتمالات ، ومن هذه الاحتمالات التي تم وضعها للبيانات مولدة وفق الأنموذج L1 موضحة بالنقاط الآتية:

- 1- عدد السكان الابتدائي $N=200$ عدد الجزر $NI=40$ وعدد الأجيال $MI=5$.
 - 2- الحد الأعلى للرتب هو $(b_0 = 3)$ و $(a_0 = 3)$
 - 3- قيد المدى الأدنى $sp(a_j + b_j) = 100 + 50(a_j + b_j)$.
 - 4- احتمال الموقع (j) من السلسلة يؤخذ كنقطة تغيير $\pi_B = sp(0) / n = 100 / 500 = 0.2$
 - 5- احتمالية إجراء عملية العبور $\pi_c = (n - sp(0)) / n = (500 - 100) / 500 = 0.8$.
 - 6- احتمالية إجراء عملية التغيير $1 - \pi_c = 0.2$.
 - 7- معيار توقف الخوارزمية الوراثية عدد الهجرات الكلية $MM=50$.
 - 8- عدد تكرارات الخوارزمية الوراثية $numbiter=100$
- الجدول الآتي يوضح بعض نتائج تطبيق الخوارزمية الوراثية على البيانات المولدة بحجم (500) مشاهدة ، وبقطع واحد عند الزمن $(k=300)$.

الجدول (1): يوضح نتائج الخوارزمية الجينية لسلسلتي المدخلات والمخرجات المولدتين وفق الأنموذج L1 البالغ عددهما 500 مشاهدة.

عدد القطاعات	عدد نقاط التغيير	النسب المئوية %
قطاع واحد	0	16%
قطاعان	1	84%
قطاعات ثلاثة	2	0%
اكثر من ثلاثة قطاعات	> 2	0%
المجموع		100%

نستنتج من خلال ملاحظة الجدول في أعلاه احتواء النظام على نقطة تغيير واحدة ، لأنه يمتلك أعلى نسبة مئوية وهي 84%. والشكل (1) يوضح مواقع نقطة التغيير التي تم الحصول عليها من خلال تطبيق الخوارزمية الجينية.



الشكل (1) مواقع نقطة التغيير للبيانات المولدة وفق الأنموذج L1 بحجم (500) مشاهدة. بملاحظة الشكل في أعلاه نجد بأنه تم ظهور نقطة التغيير عند الموقع (300) بنسبة 70% من النسبة 84% المقابلة لظهور نقطة تغيير واحدة في النظام ، وهي مطابقة لموقع نقطة التغيير التي تم توليد البيانات منها. وبذلك نستنتج بأن النظام الذي تم توليده وفق الأنموذج L1 هو نظام غير مستقر ويحتوي على نقطة تغيير واحدة وهي عند الزمن (300) والنماذج التي تم الحصول عليها ضمن كل قطاع هي: بالنسبة للقطاع الأول فإنه تم الحصول على الأنموذج $ARX(1,1,1)$ لأنه ظهر بنسبة 47% من 70% ، وهي أعلى نسبة . أما بالنسبة للقطاع الثاني تم الحصول على الأنموذج $ARX(1,2,1)$ لأنه ظهر بنسبة 36% من 70% وهي أعلى نسبة. نلاحظ بأنه تم الحصول في القطاع الثاني على أنموذج مطابق للأنموذج الذي تم توليد البيانات منه ، وهذه نتيجة جيدة وتعطي الدعم الأكبر لاستخدام الخوارزمية الجينية لإيجاد عدد ومواقع نقاط التغيير والرتب الخاصة بكل قطاع.

ثانياً: التطبيق على بيانات مولدة وفق الأنموذج L2 بعدد بيانات (1000) مشاهدة وينقضي التغيير. فقد تم توليد أنموذج $ARX(2,2,1)$ لبيانات القطاع (القسم) الأول و أنموذج $ARX(1,2,1)$ لبيانات القطاع الثاني أما القطاع الثالث فقد تم توليد أنموذج $ARX(2,2,1)$ له والصيغة العامة للأنموذج هي :

$$L2: y_t = \begin{cases} 0.2y_{t-1} - 0.4y_{t-2} + 0.2u_{t-1} - 0.3u_{t-2} + e_t & \text{if } 1 \leq t \leq 400 \\ 0.96y_{t-1} + 0.4u_{t-1} - 0.5u_{t-2} + e_t & \text{if } 401 \leq t \leq 700 \\ 0.2y_{t-1} - 0.4y_{t-2} + 0.2u_{t-1} - 0.3u_{t-2} + e_t & \text{if } 701 \leq t \leq 1000 \end{cases}$$

الاحتمالات التي تم وضعها لتطبيق الخوارزمية الجينية وفق هذه البيانات معظمها مشابهة للاحتمالات التي تم وضعها عند التطبيق على بيانات المولدة وفق الأنموذج L1 فقط تختلف في النقاط الآتية:

- 1- قيد المدى الأدنى $sp(a_j + b_j) = 100 + 40(a_j + b_j)$.
- 2- احتمال الموقع (j) من السلسلة يؤخذ كنقطة تغيير $\pi_B = sp(0)/n = 100/1000 = 0.10$
- 3- احتمالية إجراء عملية العبور $\pi_c = (n - sp(0))/n = (1000 - 100)/1000 = 0.9$.
- 4- احتمالية إجراء عملية التغيير $1 - \pi_c = 0.1$.

الجدول الآتي يوضح بعض نتائج تطبيق الخوارزمية الجينية على البيانات المولدة بحجم (1000) مشاهدة وبقطتي تغيير ، إذ القطع الأول عند الزمن ($k=400$) ، والقطع الثاني عند الزمن ($k=700$). وتم تحديد عدد القطاعات ومواقع نقاط التغيير وكذلك النماذج الخاصة ضمن كل قطاع وفق برنامج تمت كتابته باستخدام برنامج matlab2015.

الجدول (2): يوضح نتائج الخوارزمية الجينية لسلسلي المدخلات والمخرجات المولدتين وفق الأنموذج L2 البالغ عددهما 1000 مشاهدة.

عدد القطاعات	عدد نقاط التغيير	النسب المئوية %
قطاع واحد	0	0%
قطاعان	1	0%
ثلاثة قطاعات	2	88%
أربعة قطاعات	3	12%
أكثر من أربعة قطاعات	> 3	0%
المجموع		100%

نلاحظ من خلال الجدول في أعلاه النظام يحتوي على نقطتي تغيير ، لأنه يمتلك أعلى نسبة مئوية وهي 88%. والجدول الآتي يوضح مواقع نقطتي التغيير التي تم الحصول عليهما من نسبة 88%.

الجدول (3): يوضح مواقع نقطتي التغيير التي تم الحصول عليهما من خلال تنفيذ الخوارزمية الجينية للبيانات المولدة وفق الأنموذج L2 البالغ عددهما 1000 مشاهدة.

ت	مواقع نقطة التغيير الاولى	مواقع نقطة التغيير الثانية	النسب المئوية %
1	347	695	4%
2	348	697	1%
3	360	700	12%
4	373	707	1%
5	377	709	1%
6	400	700	59%
7	400	740	5%
8	401	701	4%
9	403	702	1%
	المجموع		88%

بملاحظة الجدول السابق نجد بأنه تم ظهور نقطة التغيير الأولى عند الموقع (400) ونقطة التغيير الثانية عند الموقع (700) ، فقد ظهرت هاتان النقطتان بنسبة 59% من النسبة 88% المقابلة لظهور نقطتي تغيير في النظام ونقطتي التغيير التي تم الحصول عليهما وهما مطابقتان لنقطتي التغيير التي تم توليد البيانات منها . وبذلك نستنتج بأن النظام الذي تم توليده وفق الأنموذج L2 هو نظام غير مستقر ويحتوي على نقطتي تغيير هما عند الزمن (400) و(700). والجدول الاتي يوضح رتب النماذج الخاصة بالقطاعات الثلاثة التي تم الحصول عليها من خلال تطبيق الخوارزمية الجينية.

الجدول (4): يوضح إحصائية رتب النماذج الثلاثة عند نقطتي التغيير k=400 و k=700 التي تم الحصول عليهما من خلال تنفيذ الخوارزمية الجينية.

ت	نماذج القطاع الأول	النسب المئوية %	نماذج القطاع الثاني	النسب المئوية %	نماذج القطاع الثالث	النسب المئوية %
1	ARX (2,1,7)	9%	ARX (2,3,2)	33%	ARX (1,2,1)	4%
2	ARX (2,2,7)	15%	ARX (3,2,2)	25%	ARX (1,3,1)	1%
3	ARX (2,3,7)	8%	ARX (3,3,2)	1%	ARX (2,1,1)	7%
4	ARX (3,1,7)	9%			ARX (2,2,1)	15%
5	ARX (3,2,7)	13%			ARX (2,3,1)	10%
6	ARX (3,3,7)	5%			ARX (3,1,1)	4%
7					ARX (3,2,1)	14%
8					ARX (3,3,1)	4%
	المجموع	59%		59%		59%

نلاحظ من خلال الجدول في أعلاه أنَّ الأنموذج ARX(2,2,7) الخاص بالقطاع الأول والأنموذج ARX(2,3,2) الخاص بالقطاع الثاني وكذلك الأنموذج ARX(2,2,1) الخاص بالقطاع الثالث يمتلكان أعلى نسب ، لذلك تم اعتمادها كنماذج ملائمة للقطاعات الثلاثة التي تفصلها نقطتا التغيير (k=400) و (k=700). ونلاحظ بأنَّ الأنموذج الخاص بالقطاع الثالث مطابق للأنموذج الذي تم توليد البيانات منه.

(7) الاستنتاجات والتوصيات: الاستنتاجات:

- 1- تمكنا من الوصول إلى النتائج الحقيقية، فقد تم الحصول على عدد نقاط التغيير نفسها التي تم توليد البيانات على أساسها.
- 2- حينما تم توظيف مبدأ MDL مع الخوارزمية الوراثية تمكنا من الحصول على نتائج جيدة فقد تم الحصول على مواقع وعدد نقاط التغيير مطابقة لمواقع وعدد نقاط التغيير للبيانات المولدة.
- 3- أمكن الحصول على نماذج بعملية الأنموذج الأصلي نفسه التي ولدت منه البيانات وبرتب قريبة جداً من رتب الأنموذج الأصلي باستثناء الجزء الأخير من البيانات فقد قدم أنموذجاً مطابقاً تماماً للأنموذج الأصلي الذي ولدت منه البيانات.

التوصيات:

على ضوء الاستنتاجات التي تم الحصول عليها نوصي بما يأتي:

- 1- تطبيق عملية التقطيع على النماذج الحركية التصادفية الخطية متعددة المدخلات ومتعددة المخرجات ، والبحث في إمكانية التوصل إلى نتائج جيدة.
- 2- استخدام تقنية الشبكات العصبية والمنطق المضرب في عملية تشخيص النظم الحركية الخطية غير المستقرة بتوظيف أسلوب التقطيع وصولاً لتشخيص الأنموذج الأفضل الذي يمثل البيانات قيد البحث .
- 3- من خلال النتائج التي حققت عند استخدام مبدأ MDL مع الخوارزمية الجينية يوصي الباحث باستخدام مبدأ MDL مع نماذج أخرى غير أنموذج ARX.

(8) المصادر:

1. الشرابي ، نجلاء سعد ابراهيم (2017) : " دراسة أساليب تقدير نقاط التغيير في تشخيص بعض النظم الحركية مع التطبيق " ، أطروحة دكتوراه ، كلية علوم الحاسبات الرياضيات ، جامعة الموصل ، العراق، ص41-ص44.
2. Amiya H. , Arindam K. and Soumajit P. (2012) : " Histogram based Evolutionary Dynamic Image Segmentation", International Conference on Electronics Computer Technology,P:585-589.
3. Badagian A. L. (2013) : " Time series segmentation procedures to detect, locate and estimate change-points", Universidad Carlos III de Madrid – Spain,P:1-17.
4. Cheng G. and Xiaoyong Y. (2011) : "A Programming of Genetic Algorithm in Matlab7.0", Modern Applied Science Vol. 5, No. 1,P:230-235.

5. Davis R. , Lee T., and Rodriguez G. (2006) : "Structural Break Estimation for Nonstationary Time Series Models", Journal of the American Statistical Association 101 (473), P:229–239.
6. Hamburg A. P. (2005) : "Genetic Algorithms for Optimization", Technical University Sofia,P:3-15.
7. Hamed A. , Mohammadi K. and assanpour H. (2011) : "An Improved Signal Segmentation Method using Genetic Algorithm", International Journal of Computer Applications, Volume 29– No.8,P:5-9.
8. Richard A. D. , Thomas C. M. and Gabriel A. R. (2008) : " Break Detection For A Class of Nonlinear Time Series Models", Journal of Time Series Analysis ,VOL.29,NO.5,P:835-867.
9. Rojas R. (1996) : "Genetic Algorithms", Neural Networks, Springer-Verlag, Berlin,P:429-450.