

مرشح كالمن الموسع لتدريب الشبكات العصبية الاصطناعية ذات التغذية الامامية

د.طالب شريف جليل/لثلية الادارة والاقتصاد/جامعة صلاح الدين-اربيل
سامان حسين محمود/لثلية الادارة والاقتصاد/جامعة صلاح الدين-اربيل

المستخلص

يناقش هذه البحث استخدام الشبكات العصبية الاصطناعية متعدد الطبقات ذات التغذية الامامية Feed Forward والتي يتدرب باستخدام خوارزمية مرشح كالمن الموسع مرة ، واخرى بخوارزمية الانتشار العكسي Back propagation - الطريقة الاعتيادية - ويقارن بينهما مستخدما نوعين من البيانات : هي البيانات الاصلية التي تمثل وقت تأخير لشبكة (Aria phone) للاتصالات الهاتفية والبيانات المولدة، وقد اكدت بان اسلوب مرشح كالمن اسرع لعدد خطوات الفترة الزمنية لتدريب الشبكة العصبية مقارنة بالطريقة الثانية. الكلمات الدالة: مرشح كالمن، مرشح كالمن الموسع، الشبكات العصبية

Extended Kalman Filter of Training Feed Forward Artificial Neural Network

Abstract

In this paper discuss the use of artificial neural networks multilayer feed forward, which are training using and Extended Kalman Filter (EKF) algorithm one and training Back propagation algorithm standard way then compare them by using two types of data which are the original data represent time delay of the network (Aria phone) for telephone communications and simulation data. The results confirm that EKF approach is faster to the period time steps to train the neural networks compared to the second way.

١- المقدمة

الشبكات العصبية الاصطناعية واحدة من اكثر الأساليب شيوعا للذكاء الاصطناعي ، وهي من الطرائق الحديثة التي لها الكفاءة العالية في إعطاء نتائج مرضية وجيدة في التقدير، يعود الفضل لاول شبكة عصبية اصطناعية الى العالمين (McCulloch and Pitts,1943) حيث وضعوا اول نموذج رياضي لخلية عصبية، كانت عبارة عن خلية بسيطة ثنائية الحالة (Two State Binary) وعتبة (Threshold) ثابتة

لتمثل الدوال المنطقية فيها وبأوزان (weights) متساوية، ووصفا النموذج على انه وُحد تمحسوبة لعينة خطية بمدخلات متعددة واخراج واحد. (أبو جزر وآخرون، ٢٠١٠)، (سليمان، ٢٠٠٨)

وقام الباحث (Donald Hebb, 1949) بتطوير أول قاعدة لتعلم الشبكات العصبية الاصطناعية وأطلق عليها (Hebbian Learning Rule) وقام (Frank Rosenblatt, 1959) بتطوير أول شبكة عصبية اصطناعية اطلق عليها خلية الادراك الحسي (Perceptron) وكانت من طبقة واحدة (Single layer) وكان التعليم مقتصرًا على النماذج الخطية (Linear models). (عيسى، ٢٠٠٠)

واضاف (Paul Werbos, 1974) لأول مرة الطبقات المخفية (Layers Hidden) الى الشبكات العصبية مما اكسب الادراك الحسي القدرة على حساب الدوال والعلاقات للنماذج الخطية وغير الخطية، وربط بين مستوى الادخال والمخفي ب أوزان تعدل باستخدام خوارزمية الانتشار الخلفي (العكسي) (BackPropagation)، واصبح يدرّب الشبكات العصبية متعدد الطبقات (Multilayer networks) عليها لتعلم وتوسع استخداماتها لتشمل مجالات عديدة، ومنذ ذلك الحين وحتى الان تطوّر الشبكات العصبية الاصطناعية باستمرار وتتحسن. (الكسو، ٢٠٠٥)، (Hagan & others, 1996)

وجزء هام من بناء الشبكة العصبية هو استخدام خوارزمية دقيقة وقوية في التعلم بغض النظر عن معمارية الشبكة العصبية المناسبة، والخوارزمية الأكثر شعبية في التعلم هي الشبكات العصبية الاصطناعية ذات التغذية الامامية بالتعلم تحت الاشراف، وتعتمد الشبكة العصبية ذات الانتشار العكسي للخطأ على طريقة الانحدار التدريجي (Gradient descent) وذلك لايجاد القيمة الصغرى لمربع الخطأ الكلي لقيمة المخرجات المحسوبة من قبل الشبكة وهي سهلة التطبيق وكفوءة في اغلب الاحيان، ولكن قد تكون نتائج خوارزمية الانتشار العكسي ضعيفة وغير مرضية، لذا عملية البحث عن ايجاد طرق اكثر كفاءة ولا تتأثر بالضوضاء ومن اهم هذه الطرق المستخدمة في تعلم الشبكات ال عصبية هي خوارزمية مرشح كالمن الموسع (Extended Kalman Filter) التي تقوم بتقدير اوزان الشبكة العصبية ذات التغذية الامامية والتي اقترحت الفكرة وطبقت لأول مرة في تدريب الشبكات العصبية متعدد الطبقات من قبل (Singhal and Wu, 1989) والتي تنتج نحو طريقة اقل تكرار من خوارزمية الانتشار العكسي. (Singhal & Wu, 1989)

حيث قام (Rudolf E. Kalman) بدراسة الأنظمة الديناميكية (حركية) بطريقة تعاقبية في أزمنة منقطعة والحصول على القيم التي تكون قريبت من القيم الحقيقية للقياسات باقل تباين لخطأ التقدير باستخدام

طريقة المربعات الصغرى عندما تكون الاخطاء العشوائية (الضوضاء) وأخطاء أخرى خاضعة للتوزيع الطبيعي (Gaussian)، ومن ثم قام [Kalman & Bucy, 1961] بوضع خوارزمية لتقدير الحالة الحركية في الأزمنة المستمرة (Continuous Time) وهو ما عرف لاحقاً بـ (Kalman-Bucy Filter). (Welch&bishop,2002)، (Kalman,1960)، (Grewal & Andrews,2008)

مرشح كالمن من المرشحات القوية جداً لتقدير متغير الحالة للزمن الحاضر والمستقبل اعتماداً على المعلومات الماضية حيث يقوم بتحديث (Update) المقدرات التي تتضمن في كل نقطة زمنية مشاهدة جديدة لذا لا يتطلب ذلك تخزين جميع البيانات الماضية. (Welch&bishop,2002)، (Vaseghi,2008)

والهدف من هذا البحث هو دراسة مرشح كالمن الموسع العصبية المستخدمة في تقدير اوزان الشبكة العصبية ومقارنة نتائجه مع نتائج الشبكات العصبية الامامية ذات خوارزمية الانتشار العكسي.

٢- مرشح كالمن الديناميكي (الحركي) Kalman Filter

مرشح كالمن هو موضوع واسع في البحث والتطبيق في مختلف المجالات، حيث طبق في تقدير مسار المركبة الفضائية ابولو، وفي توقع تقلبات سوق الاسهم المالية قصيرة الأمد وكذلك تستخدم في تطبيقات نظام تحديد الموقع العالمي (GPS) لتقدير الموقع والحصول على أبعاد كشف دقيقة جداً إضافة الى العديد من التطبيقات الأخرى، ويرجع جزء كبير من ذلك إلى التقدم في مجال الحاسبات الرقمية. (Vaseghi,2008)

يتكون مرشح كالمن في الحقيقة من معادلتين (معادلة المشاهدة ومعادلة الحالة) وعندما تكتب كخوارزمية يحتاج الى مجموعة من العلاقات التي تربط تقديرات الماضي والحاضر والمستقبل وهذا التقدير هو الامثل، وكشف هذا المرشح من قبل (Rudolf E. Kalman) الذي نشر بحثه المشهور في عام 1960 ووصف فيه ما توصل اليه من نتائج وهي الحلول الحسابية والكفوة لمسألة التقدير الخطي وايجاد المقدر الامثل لمتغير الحالة او المعلمة. (Kalman,1960).

٣- مرشح وخوارزمية كالمن الموسع

Extended Kalman Filter and Algorithm

يهدف مرشح كالمن إلى تقدير العملية متجه الحالة في النموذج الخطي، ولكن في النماذج غير الخطية نظريات التقدير الخطية لن تعطي نتائج جيدة وفي هذه الحالة نحتاج الى تقديرات غير خطية من خلال مرشح كالمن الموسع (EKF).

ويستخدم مرشح كالمن الموسع عندما تكون معادلة الحالة و/ او معادلة مشاهدة النظام غير الخطية ويتعامل مع المعادلات بنفس الطريقة التي يعمل مرشح كالمن الخطية ماعدا تقريب الى الخطية (linearization) لمعادلة الحالة و/ او معادلة المشاهدة حول اخر تقدير الحالة الأخيرة لكل خطوة زمنية لمرشح كالمن الموسع، ان معادلات تحديث الزمن يقرب خطياً حول تقدير حالة اللاحق (Posterior state) (\hat{x}_t) ومعادلة المشاهدة يقرب حول تقدير حالة السابق (prior state) (\hat{x}_t^-) . وتتم عملية تقريب الخطية باستخدام مفكوك سلسلة تايلر من الرتبة الأولى (First Order of Taylor Series) والرتبة الثانية. (Simon,2001) ويمكن وصف نموذج خاص غير خطي بمعادلات التالية:

$$y_t = h(x_t) + v_t \quad \text{----- (1)} \quad \text{(معادلة المشاهدة)}$$

$$x_{t+1} = f(x_t) + w_t \quad \text{----- (2)} \quad \text{(معادلة الحالة)}$$

بافتراض أن الأخطاء v_t و w_t كل واحدة منهما غير مترابطة فيما بينهما ومستقلة عن بقية المتغيرات ويتوزعان توزيعاً طبيعياً بمتوسط صفر ومصفوفات التباين V_t و W_t على التوالي. والدالة $f(x_t)$ تشير الى دالة غير خطية لمتغيرات الحالة في زمن (t) والتي من المحتمل ان تكون غير مستقرة (Time variant) والدالة $h(x_t)$ دالة غير خطية والتي من الممكن ايضاً ان تكون غير مستقرة . (Welch&bishop,2002) ، (Grewal & Andrews,2008) .

تقريب سلسلة تايلر إلى الرتبة الأولى لدوال معادلات (١ و 2) يعد من أكثر الطرق دقة والأسهل تفسيراً ويحتاج إلى افتراض أن كلا من $f(x_t), h(x_t)$ أو (على الأقل واحدة منهما) دوال قابلة للاشتقاق. (West & Harison,1997)

وان الخطوة الاولى لحساب عامل التصحيح يبدأ بتقريب معادلة المشاهدة (١) للحالة الاخيرة وهو تقدير السابق (\hat{x}_t^-) لذا يتم تقريب حول هذه الحالة باستخدام سلسلة تايلر:

$$h(x_t) \approx h(\hat{x}_t^-) + H_t^T (x_t - \hat{x}_t^-) + R_t \quad \text{----- (3)}$$

حيث أن R_t تمثل الحدود الم تبقيّة من الرتبة الثانية فما فوق ، وأن H_t تمثّل مصفوفة جاكوبيان (Jacobian) للاشتقاق الجزئية لدالة $h(\cdot)$ للحالة تقدير السابق (\hat{x}_t^-) .

$$H_t = \left. \frac{\partial h(x)}{\partial x} \right|_{x = \hat{x}_t^-} \quad \text{----- (4)}$$

وأن H_t تكون معروفة.

الحدود الثانية فما فوق في المعادلة (٣) غير خطية قد تم إهمالها لذلك تصبح معادلة المشاهدة (١) بالصيغة الآتية :

$$y_t = H_t^T x_t + b_t + v_t \quad \text{----- (5)}$$

$$b_t = h(\hat{x}_t^-) - H_t^T \hat{x}_t^- \quad \text{----- (6)} \quad \text{وان } b_t \text{ تساوي}$$

ونحسب رحية كالمن K_t باستخدام تقريب خطي لجاكوبيان

$$K_t = P_t^- H_t^T (H_t P_t^- H_t^T + V_t)^{-1} \quad \text{----- (7)}$$

والخطوة التالية في حساب تقدير مصفوفة تغاير اللاحق (posterior covariance) لمرشح كالمن الموسع

$$P_t = (I - K_t H_t) P_t^- \quad \text{----- (8)}$$

وتطبيق نفس الإجراءات على معادلة الحالة (٢)، والتي يقترب خطياً حول تقدير حالة اللاحق (\hat{x}_t^-) باستخدام سلسلة تايلر كالاتي:

$$f(x_t) \approx f(\hat{x}_t^-) + F_{t,t-1} (x_t - \hat{x}_t^-) + \delta_t \quad \text{----- (9)}$$

حيث أن δ_t تمثل الحدود المتبقية من الرتبة الثانية فما فوق وأن $F_{t,t-1}$ تمثّل مصفوفة جاكوبيان ذات بعد $(k \times k)$ لمشتقة معادلة الحالة عند التقدير \hat{x}_t^- ، وأن:

$$F_{t,t-1} = \left. \frac{\partial f(x)}{\partial x} \right|_{x=\hat{x}_t} \quad \text{----- (10)}$$

والذي يكون معروف أيضاً.

والحدود الثانية فما فوق من معادلة (٩) غير خطية قد تم إهمالها لذلك تصبح معادلة الحالة لمتغير حالة x_{t+1} تقريباً بالصيغة الآتية

$$x_{t+1} = F_{t,t-1}(x_t) + w_t + g_t \quad \text{----- (11)}$$

$$g_t = f(\hat{x}_t) - F_{t,t-1}(\hat{x}_t) \quad \text{----- (12)} \quad \text{وان}$$

$$\hat{x}_0 = E(x_0) \quad \text{----- (13a)}$$

$$P_0 = E[(x_0 - E(x_0))(x_0 - E(x_0))^T] \quad \text{----- (13b)}$$

P_0 تمثل القيمة البدائية (Initial value) لمصفوفة التغيرات P_t عندما $t=0$ اي قبل اضافة المشاهدة.

$$\hat{x}_t^- = f(\hat{x}_{t-1}^-) \quad t=1,2,\dots \quad \text{لتحديث (Propagation) لتقدير حالة الزمن}$$

٤- الشبكات العصبية الاصطناعية Artificial Neural Networks

سعيًا للوصول إلى صنع آلة أو جهاز قادر على التعلم و اكتساب المعرفة ، قام بعض العلماء بمحاولة محاكاة الجهاز العصبي و بالأخص الخلية العصبية (التي هي الوحدة الأساسية لبناء الجهاز العصبي) وتنتشر مليارات من الخلايا العصبية في جسم الانسان و تنفرع إلى زوائد تعمل على نقل الشعور بالإحساس ، ووردود الأفعال من و إلى الدماغ، ودراسة كيفية معالجة الخلية العصبية الحيوية للمعلومات وتخزينها واسترجعها ومن

ثم الاستفادة منها للتوصل الى بناء هيكلية مترابطة لبنية برمجية تشابه في ادائها للخلايا العصبية البشرية وتوصلوا إلى علم سمي بالشبكات العصبية الاصطناعية (Artificial Neural Networks) والذي يندرج تحت علوم الذكاء الصناعي، حيث يجعلون من أجهزة الحاسوب أجهزة ذكية، بإمكانها أن تكتسب المعرفة بطريقة مشابهة للطريقة التي يكتسب بها الإنسان المعرفة.

والشبكات العصبية الاصطناعية هي نظام حاسوبي يستند الى نماذج رياضية مصصمة لمحاكاة الطريقة التي يؤدي بها الدماغ البشري مهمة معينة، ومكونة من عدد كبير جدا من وحدات المعالجة (Processing Units) المترابطة مع البعض وتتصف بطبيعتها الديناميكية والمتوازية (Parallelism) في معالجة البيانات الداخلة اليها، ووحدات معالجات بسيطة ما هي الا عناصر حسابية تسمى عصبونات او عقد (Neural or Nodes) والتي لها خاصية عصبية حيوية حيث تقوم بتخزين المعرفة والمعلومات التجريبية لتجعلها متاحة للمستخدم وذلك عن طريق ضبط الاوزان، ولها خاصية توزيع الذاكرة (Distributed Memory) لذا فان فقدان احد العناصر لا يؤدي الى فقدان كامل للبيانات واما يحدث تغير هامشي على فعالية الخلايا ومن الخصائص الاخرى التي تتمتع بها الشبكة الاصطناعية هي القدرة على التكيف (Capacity Adaptation) والتنظيم الذاتي (Self-organization). (ابو جزر واخرون، ٢٠١٠)، (الكسو، ٢٠٠٥)، (Graupe, ٢٠٠٧)

5- معمارية الشبكة العصبية الاصطناعية

Artificial Neural Networks Architecture (structure)

الشبكات العصبية الاصطناعية تعمل كمجموعة متوازية من وحدات معالجة بسيطة (عقد، عصبونات)، تكتسب الاتصالات البينية بين هذه الوحدات اهمية خاصة وتقوم بدور كبير في انشاء ا لشبكة وإن ترتيب العصبونات في الطبقات أو المستويات، وشكل الترابطات ضمن أو بين الطبقات هو ما يدعى بالبنية الهندسية أو معمارية (هيكلية) الشبكة العصبية؛ ويتم حساب عدد الطبقات في الشبكة بدون حساب طبقة الإدخال لأنها لاتجرى فيها أية عملية حسابية، أو يمكن معرفتها من عدد الترابطات بين العصبونات. (عيسى، ٢٠٠٠)، (Hagan & others, 1996)

وقد تكون جميع عصبونات نفس الطبقة مرتبطة مع عصبونات طبقة اخرى او غير مرتبطة معها، مثلا عند ارتباط عصبون في الطبقة الخفية مع عصبون في طبقة الاخراج فلا بد ان تربط كل وحدة خفية مع كل عصبونات طبقة الاخراج. (Tan,2001) .

6-انواع الشبكات العصبية الاصطناعية

Types of Artificial Neural Networks Neural

ومن المميزات التي تملكها منهجية الشبكات العصبية الاصطناعية، مرونة طبقة تحديد ترابط الاوزان بين طبقة المدخلات وطبقة المخرجات فقد تكون شبكات ذات طبقة واحدة Single Layer Networks او شبكات متعدد الطبقات Multilayer Networks MLN وتتكون هذه الشبكات من ثلاث طبقات ، طبقة الادخال، وطبقة الاخراج، ويتوسطهما الطبقات المخفية (طبقة واحدة او أكثر) عدد المدخلات وعدد الخلايا (العقد) في الطبقة المخفية تحدد بحسب درجة تعقيد المسألة وحجم الإدخال إلى أن نصل إلى النموذج الأمثل. (الكسو، ٢٠٠٥)، (Heaton,2005) .

٧-تدريب (او التعلم) الشبكات العصبية الاصطناعية

Training Artificial Neural Networks

تعد عملية التعلم اساساً لاداء اي عملية من عمليات الشبكات العصبية، والتعلم هي عملية تكيفية للشبكة من خلال ضبط الاوزان المرتبطة بكل الخلايا، حيث أن الشبكات العصبونية لا تيرمج بل تقوم بالتعلم ويجري التعليم عن طريق خوارزميات متخصصة تتولى تغيير قوة الربط الداخلي (الاوزان) وهذا ما يدعى بتكييف الوزن والذي يحدث خلال عملية تدعى مرحلة التدريب ويتم اختيار الاوزان الاولية (Initial) للتدريب بشكل عشوائي او اعتماداً على بعض التجارب السابقة. (الكسو، ٢٠٠٥)، (Simon,1999) .

٨-الشبكات ذات التغذية الامامية والتعلم بالاشراف

Feed Forward Neural Network and Supervised Learning

وفيها يكون اتجاه انتشار الاشارات الداخلة في الشبكة الى الامام دائماً، وبذلك فان الاشارة الخارجة من أي خلية تعتمد على الاشارات الداخلة فقط، ويتعلم من خلال التعليم بالاشراف (Supervised Learning) حيث في هذه الطريقة تحتاج الشبكات العصبية الاصطناعية الى وجود زوجين من المتجهات ، هما متجه

الإدخال ومنتج الإخراج المطلوب (Desired Output)، تدرب الشبكة العصبية على عدد من المنتجات لهذه الأزواج، تبدأ عملية التدريب بمنتج الإدخال حيث تطبق على الشبكة فينتج الإخراج الحقيقي (Actual output) ويقارن مع ما يقابله من منتج الإخراج المتوقع والفرق بينهم يمثل الخطأ (Error) الذي يستخدم لتعديل الأوزان طبقاً لخوارزمية التعليم، ويطبق جميع المنتجات الى ان يصل الخطأ لمجموعة التدريب المدخلة إلى اقل ما يمكن، ويسمى أيضاً التعليم بالإشراف اوالتعلم مع المعلم (Learning with Teacher). (Simon,1999).

٩- الشبكات العصبية ذات الانتشار العكسي Backpropagation Neural Network

تعد من اهم الشبكات العصبية الاصطناعية وأكثرها استخداما في التعلم، التي تستخدم في الشبكات العصبية متعدد الطبقات (طبقة خفية واحدة او اكثر) ذات التغذية الامامية بالتعلم تحت إشراف، تعتمد الشبكة العصبية ذات الانتشار العكسي للخطأ على طريقة الانحدار التدريجي (Gradient descent) وذلك لاجاد القيمة الصغرى لمربع الخطأ الكلي لقيمة المخرجات المحسوبة من قبل الشبكة حيث تحدث الأوزان بين الطبقات لحين الوصول إلى الأوزان المثلى التي تعطي اقل مجموع مربع خطأ ممكن بين مخرجات الشبكة العصبية وبيانات النموذج، وتسمى أيضا بخوارزمية قاعدة دلتا العمومية (العامة) Generalized Delta Rule. (الكسو، ٢٠٠٥).

١٠- مرشح كالمن الموسع والشبكات العصبية الاصطناعية

هناك طرق عديدة استعملت لتدريب الشبكات العصبية للحصول على الامثلية ومن هذه الطرق المهمة طريقة مرشح كالمن الموسع لتدريب اوزان الشبكة العصبية متعدد الطبقات والتي اقترحت الفكرة من قبل (Singhal and Wu,1989) وطور الفكرة ووضع الاساس لها من قبل (Stubberud,1990). (Singhal and Wu,1989)، (Lary & Mussa,2004)

وقد ذكرنا سابقا بان مرشح كالمن الخطي يحاول تقدير حالة النظام الخطي وتقدير معادلة المشاهدة مضافا اليهما الضوضاء ولهما توزيع كاوسياً (white Gaussian noise)، وبما ان اغلب الشبكات العصبية متعدد الطبقات نظام غير خطي لذا لا يمكن استخدام مرشح كالمن الخطي، وللحصول على تقدير امثل نقوم بعملية تقريب خطي للانظمة غير الخطية ثم يمكن تطبيق خوارزمية مرشح كالمن عليها، وان المرشح الناتج يهيى مرشح كالمن الموسع.

تستخدم خوارزمية مرشح كالمن الموسع العصبية نظام ديناميكي لتقدير حالة مرشح كالمن الموسع في نفس الوقت يقوم بتدريب اوزان الشبكة العصبية متعدد الطبقات (طبقة مخفية واحدة) ذات التغذية الامامية وكخطوة اولية نبدأ بوصف المعادلات التي تستخدم خوارزمية مرشح كالمن الموسع لتدريب اوزان الشبكة العصبية من قبل نظام غير خطي متقطع كالتالي:

$$\mathbf{w}_{t+1} = \mathbf{w}_t + \boldsymbol{\omega}_t \quad \dots\dots\dots 14$$

$$\mathbf{y}_t = \mathbf{h}(x_t, \mathbf{w}_t) + \mathbf{v}_t \quad \dots\dots\dots 15$$

الدالة الاولى تمثل معادلة الحالة التي تنظم المعلمات (الاوزان W) للشبكة العصبية متعدد الطبقات كموجة الحالة المستقرة مضافا اليها ضوضاء ω_t والدالة الثانية تمثل معادلة القياس او المشاهدة لمرشح كالمن الموسع وذلك لان دالة التحويل (Transfer function) $h(x_t, w_t)$ للشبكة العصبية الاصطناعية تكون غير خطية، وقد استخدم في هذا البحث دالة السيكمويد (Sigmoid Function) مضافا اليها الضوضاء \mathbf{v}_t و نفترض أن أخطأ \mathbf{v}_t و $\boldsymbol{\omega}_t$ كل واحدة منهما غير مترابطة فيما بينهما ومستقلة عن بقية المتغيرات ويتوزعان توزيعاً طبيعياً بمتوسط صفر ومصنوفة تباين $W_{P \times P}(t)$ و $V_{M \times M}(t)$. (Simon,2001) .

١١- تدريب مرشح كالمن الموسع لتقدير الاوزان

تدريب (تقدير الحالة) لمشكلة معينة اي تحديد اوزان (متجه الحالة) باستخدام مرشح كالمن الموسع يمكن ان يوصف من خلال ايجاد اقل متوسط مربع خطأ للتقدير minimum mean-squared error estimate تستخدم جميع المشاهدات الممكنة، حيث يفترض معمارية الشبكة بوجود M من الاوزان و N من عقدة المخرجات، والتدريب لحصول الحل للمشكلة بطريقة EKF تعتبر طريقة تكرارية وحسب الخوارزمية التالية: (Simon,2001)، (Lary & Mussa,2004) .

$$\begin{aligned} \hat{w}_t^- &= \hat{w}_{t-1} & \dots\dots\dots 16 \\ P_t^-(w) &= P_{t-1}(w) & \dots\dots\dots 17 \\ K_t &= P_t^-(w) H_t^T [H_t P_t^-(w) H_t^T + \sigma_v^2]^{-1} & \dots\dots\dots 18 \\ P_t(w) &= (I - K_t H_t) P_t^-(w) & \dots\dots\dots 19 \\ H_t &= \left. \frac{\partial h(w)}{\partial w} \right|_{w = \hat{w}_t^-} & \dots\dots\dots 20 \\ \hat{w}_t &= \hat{w}_t^- + K_t [y_t - H_t \hat{w}_t^-] & \dots\dots\dots 21 \end{aligned}$$

١٢- التطبيق

لغرض تطبيق طريقة مرشح كالمن الموسع لشبكة العصبية ذات التغذية الامامية متعدد الطبقات بلدخال واحد وعقدتين مخفيتين مستخدما تابع تنشيط (sigmoid) ومخرج واحد بتابع تنشيط خطي وان اوزان الشبكة تكون على شكل متجه التي تحدد القيم الاولية لها بشكل عشوائي ويتجدد حسب خوارزمية مرشح كالمن وتكون الغاية من تدريب الشبكة ايجاد اوزان النموذج التي تعلم (تعديل الاوزان والانحراف) برمجيا حسب اختلاف بين اخراج وادخال الشبكة لتصغير متوسط الخطأ التربيعي ، و ثم إجراء المقارنة مع طريقة الشبكات العصبية ذات التغذية الامامية مستخدماً الانتشار العكسي لتعديل الاوزان بنفس المدخلات والمخرجات والعقد المخفية واختيار نفس الدوال المذكور هانفاً.

قام الباحثان أولاً: التطبيق العملي على بيانات حقيقية تمثل بيانات عن وقت التأخير (delay Time) لشبكة (Airefon) الاتصالات الهاتفية ثانياً- ، إجراء المحاكاة لبيانات المولدة عشوائياً (Simulation)، وذلك من خلال الاعتماد على لغة ماتلاب (Matlab) في تصميم البرمجة المخصصة لهذا التحليل.

أولاً-البيانات الحقيقية لشبكة الاتصالات هاتفية (Airefon):

تم اخذ البيانات الحقيقية من مشروع Eagle (ناريا فون) وتستند على تكنولوجيا CDMA وهذه البيانات تمثل وقت تأخير وصول الإشارة لهاتف ناريا فون من هاتف اخر مبروطة معها وتم احتساب الوسط الحسابي (٢٠٢٤٩) والتباين (٠.٥٢٩)، وان الجدول (٢ و ٣) يعرض نتائج القيمة الحقيقية والقيمة التقديرية والفرق بينهما وكذلك اظهر نفس النتائج من خلال رسم بياني كما في شكل (٣،٤). وحساب بعض المقاييس ووضع افضل النتائج في الجدول (١) وحيث نلاحظ بان قيمة MSE لطريقة مرشح كالمن الموسع العصبية

للبيانات الحقيقية تساوي $3.32E-11$ وتكون اقل مقارنة بقيمة MSE لطريقة الشبكات العصبية (الانتشار العكسي) التي تساوي $5.061E-10$ وذلك بعد عديد من المحاولات.

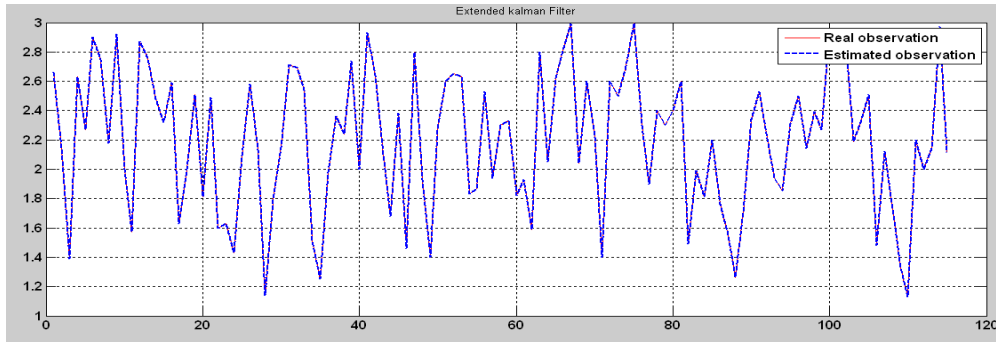
ونائج الطريقة الثانية يمكن ملاحظتها ايضاً من خلال شكل (٥-ب) الذي يبين عدد التكرارات وعملية التدريب والتحقق والاختبار وفضل قيمة MSE التي تكون عند ٣٣٥ تكرار وتكون كبيرة مقارنة بعدد التكرارات مرشح كالمن الموسع العصبية التي تساوي ٥ تكرارات. ولمقارنة قيمة الخطأ بين القيم الحقيقية القيم والمقدرة للطريقتين من خلال مقياسين اخرين MAPE و MAE كما في الجدول (١) نستنتج بان الطريقة الاولى ادق من الطريقة الثانية
ثانياً-البيانات المولدة:

للحصول على البيانات المولدة تم الاعتماد على قيمة الوسط الحسابي وقيمة التباين للبيانات الحقيقية لاشارات هاتف ثاريا فون والتي تتوزع توزيعاً طبيعياً ، وان الجدول (٢،٣) يبين البيانات المولدة والبيانات المقدره بطريقة مرشح كالمن الموسع العصبية ، ومرة اخرى بطريقة الشبكات العصبية (الانتشار العكسي) وحساب بعض مقاييس الاحصائية ووضع افضل النتائج في الجدول (١) وحيث نلاحظ بان قيمة MSE لطريقة مرشح كالمن الموسع العصبية تساوي $7.52E-12$ وتكون اقل مقارنة بقيمة MSE لطريقة الشبكات العصبية (الانتشار العكسي) التي تساوي $1.18E-10$ ولكن بعد عديد من المحاولات، وان نتائج الطريقة الاخيرة يمكن ملاحظتها من خلال شكل (٥-أ) الذي يبين عدد التكرارات وعملية التدريب والتحقق والاختبار وفضل قيمة MSE التي تكون عند ١٦١ تكراراً وتكون كبيرة مقارنة بعدد تكرارات مرشح كالمن الموسع العصبية التي تساوي ٦ تكرار . ولمقارنة قيمة الخطأ بين القيم الحقيقية القيم المقدره لطريقتين من خلال مقياسين اخرين MAPE و MAE كما في الجدول (١) نستنتج بان الطريقة الاولى ادق من الطريقة الثانية.

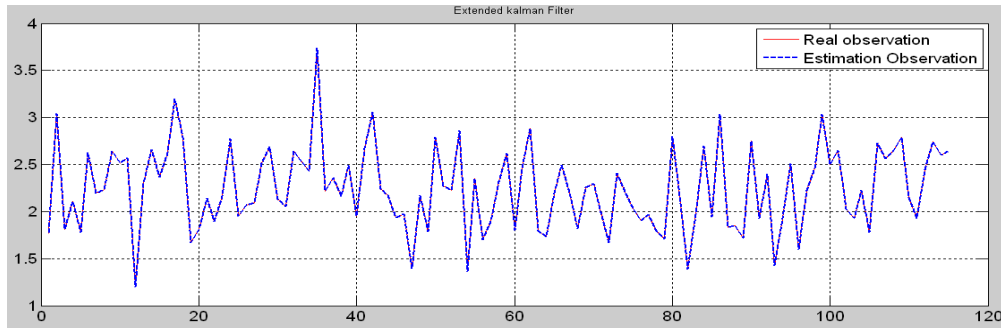
جدول (١) مقارنة بين مرشح كالمن الموسع العصبية و الشبكة العصبية ذات التغذية الامامية بالانتشار العكسي

نوع البيانات	مقاييس	مرشح كالمن الموسع العصبية	الشبكة العصبية (الانتشار العكسي)
البيانات الحقيقية	MSE	$3.32E-11$	$5.061E-10$
	MAPE	$7.8275e-05$	$1.13E-03$
	MAE	$2.07e-06$	$2.43E-5$
البيانات المولدة	التكرار(Iteration epoch)	5	335
	MSE	$7.52E-12$	$1.18E-10$
	MAPE	$5.02E-05$	$1.26E-02$
	MAE	$8.26E-07$	$2.69E-04$

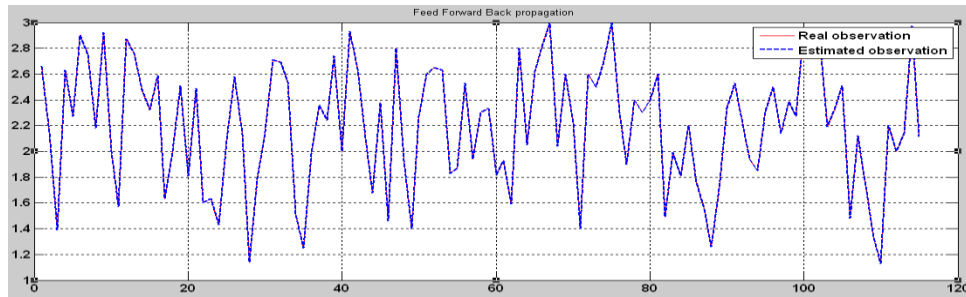
161	6	التكرار (epoch) Iteration	
-----	---	---------------------------	--



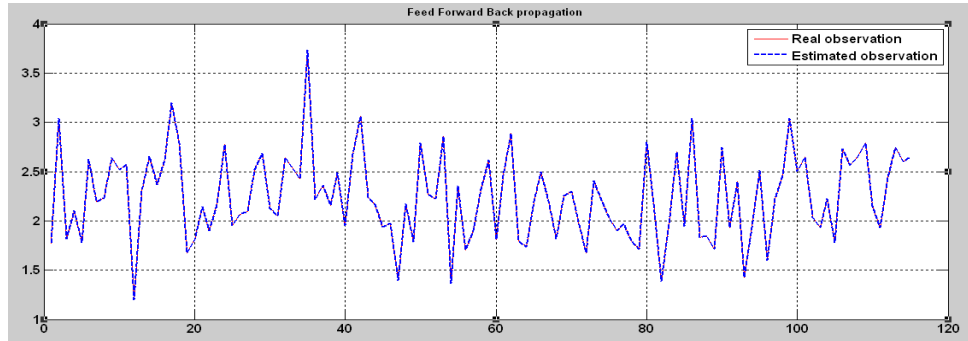
شكل (١) يمثل البيانات المولدة والبيانات المقدرة بطريقة مرشح كالمن الموسع العصبية.



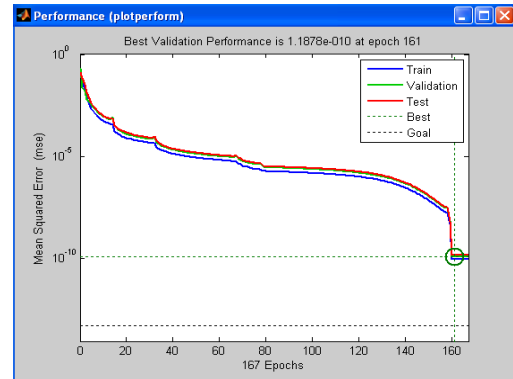
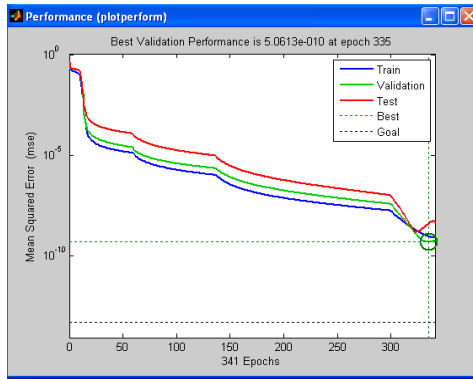
شكل (٢) يمثل البيانات الحقيقية والبيانات المقدرة بطريقة مرشح كالمن الموسع العصبية.



شكل (٣) يمثل البيانات المولدة والبيانات المقدرة بطريقة الشبكات العصبية ذات التغذية الامامية مستخدماً الانتشار العكسي.



شكل (٤) يمثل البيانات الاصلية والبيانات المقدرة بطريقة الشبكات العصبية ذات التغذية الامامية مستخدماً الانتشار العكسي.



شكل (٥) يمثل افضل تقدير بطريقة الشبكات العصبية ذات التغذية الامامية أ- الجزء الأيمن يمثل البيانات المولدة

ب- الجزء الايسر يمثل البيانات الاصلية

١٣- الاستنتاج:

نستنتج على ضوء نتائج التجريبتين اللتين استخدم فيهما نوعين من البيانات بان طريقة مرشح كالمن الموسع العصبية طريقة سريعة وتقلل عدد الخطوات الزمنية المطلوبة لتدريب الشبكة العصبية الاصطناعية مقارنة بطريقة الشبكات العصبية ذات التغذية الامامية مستخدماً الانتشار العكسي لتعديل الاوزان، حيث ح صلنا بخمس و ست تكرارات بطريقة كالمن على نفس النتائج وادق ها مقارنة بثلاث مقاييس : MSE، MAPE و MAE مع نتائج الطريقة الاخرى والذي حصل على تلك النتيجة

بـ(335،161) تكراراً أو (epoch)، وهذا يدل على سرعة مرشح كالمن الموسع العصبية ودقته مقارنة بطريقة الشبكات العصبية (الانتشار العكسي).

جدول (٢) يمثل البيانات المولدة والبيانات الاصلية مع تقديرها بطريقة مرشح كالمن الموسع العصبية

بيانات مولدة	بيانات مقطرة	بيانات مولدة	بيانات مقطرة	بيانات اصلية	بيانات مقطرة	بيانات اصلية	بيانات مقطرة
1.77	1.769	1.81	1.807	2.7	2.66	1.8	1.82
3.04	3.04	2.46	2.46	2.1	2.14	1.9	1.93
1.81	1.814	2.89	2.885	1.4	1.39	1.6	1.59
2.11	2.108	1.8	1.798	2.6	2.63	2.8	2.80
1.78	1.782	1.73	1.734	2.3	2.27	2.1	2.05
2.63	2.628	2.19	2.195	2.9	2.90	2.6	2.62
2.19	2.194	2.49	2.495	2.8	2.75	2.8	2.81
2.23	2.234	2.2	2.201	2.2	2.18	3	2.99
2.64	2.642	1.82	1.816	2.9	2.92	2	2.04
2.52	2.517	2.25	2.254	2	2.00	2.6	2.60
2.57	2.569	2.3	2.297	1.6	1.57	2.2	2.22
1.2	1.198	1.99	1.991	2.9	2.87	1.4	1.40
2.3	2.303	1.67	1.673	2.8	2.76	2.6	2.60
2.65	2.654	2.4	2.405	2.5	2.48	2.5	2.50
2.36	2.364	2.21	2.208	2.3	2.32	2.7	2.70
2.62	2.624	2.03	2.029	2.6	2.59	3	3.00
3.2	3.197	1.9	1.903	1.6	1.63	2.3	2.30
2.79	2.787	1.96	1.965	2	2.00	1.9	1.90
1.67	1.672	1.79	1.792	2.5	2.51	2.4	2.40
1.81	1.81	1.71	1.711	1.8	1.81	2.3	2.30
2.14	2.143	2.79	2.795	2.5	2.49	2.4	2.40
1.9	1.898	2.09	2.087	1.6	1.60	2.6	2.60
2.16	2.163	1.39	1.387	1.6	1.63	1.5	1.49
2.78	2.777	1.97	1.971	1.4	1.43	2	1.99
1.95	1.95	2.7	2.698	2.1	2.09	1.8	1.81
2.07	2.066	1.95	1.947	2.6	2.58	2.2	2.20
2.09	2.093	3.04	3.035	2.1	2.13	1.8	1.76
2.52	2.517	1.83	1.834	1.1	1.14	1.6	1.55
2.69	2.687	1.85	1.846	1.8	1.79	1.3	1.26
2.13	2.13	1.71	1.714	2.2	2.15	1.7	1.71
2.05	2.051	2.75	2.747	2.7	2.71	2.3	2.34
2.64	2.639	1.93	1.927	2.7	2.69	2.5	2.53
2.54	2.54	2.4	2.396	2.5	2.54	2.2	2.21
2.43	2.429	1.43	1.425	1.5	1.51	1.9	1.94
3.74	3.735	1.93	1.931	1.3	1.25	1.9	1.85
2.21	2.215	2.51	2.51	2	1.98	2.3	2.31
2.36	2.36	1.6	1.598	2.4	2.36	2.5	2.50
2.16	2.161	2.22	2.224	2.2	2.24	2.1	2.14
2.49	2.49	2.46	2.458	2.7	2.74	2.4	2.39
1.95	1.953	3.04	3.036	2	2.00	2.3	2.27

41	2.67	2.674		100	2.5	2.501	41	2.9	2.93	100	3	2.95
42	3.06	3.06		101	2.64	2.645	42	2.6	2.64	101	3	2.96
43	2.24	2.24		102	2.03	2.034	43	2.1	2.10	102	2.9	2.90
44	2.17	2.168		103	1.93	1.928	44	1.7	1.68	103	2.2	2.19
45	1.94	1.935		104	2.23	2.227	45	2.4	2.38	104	2.3	2.32
46	1.98	1.977		105	1.78	1.78	46	1.5	1.46	105	2.5	2.51
47	1.4	1.399		106	2.73	2.727	47	2.8	2.80	106	1.5	1.48
48	2.17	2.175		107	2.56	2.565	48	1.9	1.93	107	2.1	2.12
49	1.79	1.786		108	2.65	2.652	49	1.4	1.40	108	1.7	1.73
50	2.79	2.79		109	2.79	2.793	50	2.3	2.27	109	1.3	1.33
51	2.27	2.273		110	2.15	2.149	51	2.6	2.60	110	1.1	1.13
52	2.23	2.226		111	1.93	1.929	52	2.7	2.65	111	2.2	2.20
53	2.86	2.857		112	2.45	2.447	53	2.6	2.63	112	2	2.00
54	1.37	1.365		113	2.74	2.741	54	1.8	1.83	113	2.1	2.14
55	2.35	2.355		114	2.6	2.603	55	1.9	1.87	114	3	2.97
56	1.7	1.703		115	2.64	2.644	56	2.5	2.53	115	2.1	2.11
57	1.9	1.898					57	1.9	1.94			
58	2.32	2.321					58	2.3	2.30			
59	2.62	2.619					59	2.3	2.33			

جدول (٣) يمثل البيانات الاصلية والمولدة مع تقديرها بطريقة الشبكات العصبية (الانتشار العكسي)

بيانات مولدة	بيانات مقطرة	بيانات مولدة	بيانات مقطرة	بيانات اصلية	بيانات مقطرة	بيانات اصلية	بيانات مقطرة				
1	1.77	1.769	60	1.81	1.807	1	2.7	2.66	60	1.8	1.82
2	3.04	3.04	61	2.46	2.46	2	2.1	2.14	61	1.9	1.93
3	1.81	1.814	62	2.89	2.885	3	1.4	1.39	62	1.6	1.59
4	2.11	2.108	63	1.8	1.798	4	2.6	2.63	63	2.8	2.8
5	1.78	1.782	64	1.73	1.734	5	2.3	2.27	64	2.1	2.05
6	2.63	2.628	65	2.19	2.195	6	2.9	2.9	65	2.6	2.62
7	2.19	2.194	66	2.49	2.495	7	2.8	2.75	66	2.8	2.81
8	2.23	2.234	67	2.2	2.201	8	2.2	2.18	67	3	2.99
9	2.64	2.641	68	1.82	1.816	9	2.9	2.92	68	2	2.04
10	2.52	2.517	69	2.25	2.254	10	2	2	69	2.6	2.6
11	2.57	2.569	70	2.3	2.297	11	1.6	1.57	70	2.2	2.22
12	1.2	1.198	71	1.99	1.991	12	2.9	2.87	71	1.4	1.4
13	2.3	2.303	72	1.67	1.673	13	2.8	2.76	72	2.6	2.6
14	2.65	2.654	73	2.4	2.405	14	2.5	2.48	73	2.5	2.5
15	2.36	2.364	74	2.21	2.208	15	2.3	2.32	74	2.7	2.7
16	2.62	2.624	75	2.03	2.029	16	2.6	2.59	75	3	3
17	3.2	3.197	76	1.9	1.903	17	1.6	1.63	76	2.3	2.3
18	2.79	2.787	77	1.96	1.965	18	2	2	77	1.9	1.9
19	1.67	1.672	78	1.79	1.792	19	2.5	2.51	78	2.4	2.4
20	1.81	1.81	79	1.71	1.711	20	1.8	1.81	79	2.3	2.3
21	2.14	2.143	80	2.79	2.795	21	2.5	2.49	80	2.4	2.4

22	1.9	1.898		81	2.09	2.087
23	2.16	2.163		82	1.39	1.387
24	2.78	2.777		83	1.97	1.971
25	1.95	1.95		84	2.7	2.698
26	2.07	2.066		85	1.95	1.947
27	2.09	2.093		86	3.04	3.035
28	2.52	2.517		87	1.83	1.834
29	2.69	2.687		88	1.85	1.846
30	2.13	2.13		89	1.71	1.714
31	2.05	2.051		90	2.75	2.747
32	2.64	2.639		91	1.93	1.927
33	2.54	2.54		92	2.4	2.396
34	2.43	2.429		93	1.43	1.425
35	3.74	3.735		94	1.93	1.931
36	2.21	2.215		95	2.51	2.51
37	2.36	2.36		96	1.6	1.598
38	2.16	2.161		97	2.22	2.224
39	2.49	2.49		98	2.46	2.458
40	1.95	1.953		99	3.04	3.036
41	2.67	2.674		100	2.5	2.501
42	3.06	3.06		101	2.64	2.645
43	2.24	2.24		102	2.03	2.034
44	2.17	2.168		103	1.93	1.928
45	1.94	1.935		104	2.23	2.227
46	1.98	1.977		105	1.78	1.78
47	1.4	1.399		106	2.73	2.727
48	2.17	2.175		107	2.56	2.565
49	1.79	1.786		108	2.65	2.652
50	2.79	2.79		109	2.79	2.793
51	2.27	2.273		110	2.15	2.149
52	2.23	2.226		111	1.93	1.929
53	2.86	2.857		112	2.45	2.447
54	1.37	1.365		113	2.74	2.741
55	2.35	2.355		114	2.6	2.603
56	1.7	1.703		115	2.64	2.644
57	1.9	1.898				
58	2.32	2.321				
59	2.62	2.619				
22	1.6	1.6		81	2.6	2.6
23	1.6	1.63		82	1.5	1.49
24	1.4	1.43		83	2	1.99
25	2.1	2.09		84	1.8	1.81
26	2.6	2.58		85	2.2	2.2
27	2.1	2.13		86	1.8	1.76
28	1.1	1.14		87	1.6	1.55
29	1.8	1.79		88	1.3	1.26
30	2.2	2.15		89	1.7	1.71
31	2.7	2.71		90	2.3	2.34
32	2.7	2.69		91	2.5	2.53
33	2.5	2.54		92	2.2	2.21
34	1.5	1.51		93	1.9	1.94
35	1.3	1.25		94	1.9	1.85
36	2	1.98		95	2.3	2.31
37	2.4	2.36		96	2.5	2.5
38	2.2	2.24		97	2.1	2.14
39	2.7	2.74		98	2.4	2.39
40	2	2		99	2.3	2.27
41	2.9	2.93		100	3	2.95
42	2.6	2.64		101	3	2.96
43	2.1	2.1		102	2.9	2.9
44	1.7	1.68		103	2.2	2.19
45	2.4	2.38		104	2.3	2.32
46	1.5	1.46		105	2.5	2.51
47	2.8	2.8		106	1.5	1.48
48	1.9	1.93		107	2.1	2.12
49	1.4	1.4		108	1.7	1.73
50	2.3	2.27		109	1.3	1.33
51	2.6	2.6		110	1.1	1.13
52	2.7	2.65		111	2.2	2.2
53	2.6	2.63		112	2	2
54	1.8	1.83		113	2.1	2.11
55	1.9	1.87		114	3	2.97
56	2.5	2.53		115	2.1	2.11
57	1.9	1.94				
58	2.3	2.3				
59	2.3	2.33				

المصادر

- ١- ابو جزر، امجد عباس، عقل، مصباح جمعة، القاضي، زياد عبدالكريم، ابو زلطة، محمد خليل، (٢٠١٠): "الشبكات العصبية والمنطق المشوش"، طبعة الاولى، دار الاعصار العلمي للنشر والتوزيع، الاردن.
- ٢- الكسو، ابتهاج عبد الحميد محمد علي (٢٠٠٥) : " استخدام الشبكات العصبية في تقدير رتب سلاسل ماركوف مع التطبيق على سلسلة جبل بطمة في محافظة نينوى"، اطروحة دكتوراه، غير منشورة، كلية علوم الحاسبات والرياضيات، جامعة الموصل، العراق.
- ٣- سليمان، انتصار ابراهيم الياس. (٢٠٠٨): "التكهن بواسطة نماذج السلاسل الزمنية والشبكات العصبية الاصطناعية ذات التغذية الامامية مع التطبيق - دراسة مقارنة"، رسالة ماجستير، غير منشورة، كلية علوم الحاسبات والرياضيات، جامعة الموصل، العراق.
- ٤- عيسى، علام زكي (٢٠٠٠): "الشبكات العصبية، البنية الهندسية- الخوارزميات- التطبيقات"، الطبعة الاولى، شعاع للنشر والعلوم، حلب، سوريا.
- 5-Graupe, Daniel, (2007):"Principles of Artificial Neural Networks", Second edition, World Scientific Publishing Co. Singapore.
- 6-Grewal M. S. and Andrews A. P.(2008):"Kalman Filtering Theory and Practice Using MATLAB", Third Edition, John Wiley & Sons,USA.
- 7-Hagan, Martin T., Demuth, Howard B., Beale, Mark, (1996):"Neural network Design", Pws Publishing Co., USA .
- 8-Heaton, Jeff, (2005):" Programming Neural Networks in Java",Heaton Research, Inc., USA.
- 9-Kalman, R.E., (1960), "A New Approach to Linear Filtering and Prediction Problems", Journal of Basic Engineering, Vol. 82, pp. 34-45.
- 10-Lary, D.J. and Mussa, H.Y.(2004):" Using an extended Kalman filter learning algorithm for feed-forward neural networks to describe tracer correlations", Atmos. Chem. Phys. Discuss., 4, pp 3653–3667.
- 11-Simon, Haykin, (1999):" Neural Networks A Comprehensive Foundation", Second Edition, Prentice Hall International, USA.
- 12-Simon, Haykin, (2001):" Kalman filtering and neural networks ", first Edition, John Wiley & Sons,UK.

- 13-Singhal, S. and Wu, L.,(1989):”Training feed forward networks with the extended Kalman algorithm”, Proc. Int. Conf. ASSP, pp. 1187-1190.
- 14-Tan, Clarence N. W. (2001) “An Artificial Neural Networks Primer with Financial Applications Examples in Financial Distress Predictions and Foreign Exchange Hybrid Trading System “,Wilberto Publishing, Australia.
- 15-Vaseghi, S., (2008)”Advanced Digital Signal Processing and Noise Reduction”, Fourth edition, John Wiley & Sons,UK.
- 16-Welch, G. and Bishop, G., (2002), “An Introduction to the Kalman Filter”, www.cs.unc.edu/~Welch.
- 17-West, M. and Harrison, P.J., (1997),” Bayesian Forecasting and Dynamic Model”, Second edition, Springer-Verlag, New York.