

تحلیل باکس - جنکینز و شبکه های عصبی در مدل بندی بارش ایستگاه شیراز

حمیدرضا خورشیدی^۱، امین شیروانی^۲

چکیده

مدل های سری های زمانی و شبکه های عصبی مصنوعی روش های شناخته شده ای در مدل بندی داده های سری زمانی می باشند. در این مقاله ترکیب این دو روش در مدل بندی داده های بارش ماهانه مورد ارزیابی قرار گرفت. در روش ترکیبی، متغیرهای ورودی طراحی شبکه عصبی بر اساس تحلیل باکس - جنکینز بدست می آیند. این روش برای ایستگاه سینوپتیک شیراز انجام گردید. کل دوره زمانی مورد مطالعه شامل ۵۵ (۲۰۰۵-۱۹۵۱) سال بود که ۵۰ سال اولیه برای آموزش و بقیه برای آزمون مدل انتخاب گردید. با استفاده از مدل ارائه شده، مقادیر آتی ماهانه بارش با دقت بالایی پیش بینی گردیدند.

کلید واژه: بارش، پیش بینی، شبکه های عصبی مصنوعی، مدل های سری های زمانی، شیراز.

مقدمه

وجود یک مدل مناسب برای بارش هر منطقه را می توان از سه جهت مهم دانست: ۱- شناخت وضع موجود بارندگی منطقه ۲- مقایسه آن با مناطق دیگر ۳- پیش بینی که مهمترین هدف در این زمینه است. امروزه روش باکس - جنکینز [۱] متداولترین روش در مدل بندی برای سری های زمانی با مؤلفه فصلی است که در قالب الگوهای اتورگرسیو - میانگین متحرک تلفیق شده فصلی (SARIMA^۲) خلاصه می شود. سلطانی و همکاران [۶] الگوهای SARIMA را برای پیش بینی بارش در کشور استفاده کرده اند. شریفان و قهرمان [۴]

۱- مربی دانشگاه آزاد اسلامی واحد سروستان (khorshidi۷۳@yahoo.com)

۲- بخش مهندسی آب دانشگاه شیراز (am_shirvani@hotmail.com)

۳- Seasonal Auto-Regressive Integrated Moving Average

نیز با استفاده از این الگوها تخمین بارندگی ۱۰ روزه در استان گلستان را مورد ارزیابی قرار داده اند. از طرفی شبکه های عصبی مصنوعی نیز بعنوان یک رهیافت موازی با روش باکس-جنکینز در مدل بندی مشاهداتی با مؤلفه خطا بطور وسیعی بکار گرفته شده اند. از جمله کارهای مهم در این زمینه می توان به مقاله فاراوی و چتفیلد [۳] اشاره کرد که در آن داده های خطوط هوایی مطالعه گردید و عملکرد بهتر شبکه های عصبی نسبت به روش باکس-جنکینز تایید شد. در سالهای اخیر مقالاتی منتشر شده است که ترکیبی از هر دو رهیافت باکس-جنکینز و شبکه های عصبی را بکار برده اند. در این راستا بوهامرا و همکاران [۲] از روش ترکیبی در مدل بندی سری زمانی میزان مصرف آب کویت استفاده کردند. در روش ترکیبی تأخیرهای ظاهر شده در یک مدل مناسب SARIMA، بعنوان بهترین پیش بینی کننده ها، ماتریس ورودی را در طراحی یک شبکه عصبی تشکیل می دهند. این مقاله به شناسایی روش ترکیبی پرداخته و به صورت موردی نتایج آن را برای دستیابی به الگوی پیش بینی بارش ایستگاه سینوپتیک شیراز بکار برده است.

داده ها و روش ها

داده های بارش ماهانه برای ایستگاه سینوپتیک شیراز از سالنامه هواشناسی انتشارات سازمان هواشناسی کشور استخراج گردید. کل دوره آماری این ایستگاه از ۱۹۵۱ تا ۲۰۰۵ بود.

۱-۲ رهیافت باکس-جنکینز

رهیافت باکس-جنکینز شامل تبدیل مشاهدات به حالت ایستا بر اساس تفاضل های فصلی و غیر فصلی و در نهایت ایجاد یک رابطه خطی پیش بینی کننده برای مشاهدات است. با فرض اینکه $\{X_t, t \in T\}$ دنباله ای از متغیرهای تصادفی بر حسب زمان باشد، تابع میانگین و تابع اتوکواریانس آن به ترتیب با μ_t و $\gamma_{t,s}$ بیان شده و به صورت زیر تعریف می

شوند:

$$E(X_t) = \mu_t, \quad \gamma_{t,s} = \text{cov}(X_t, X_s) \quad ; \quad t, s \in T$$

^۴-Lag

در حالتی که میانگین ثابت بوده و به زمان بستگی نداشته باشد و تابع اتوکواریانس نیز بجای هر زوج زمان (t, s) تنها به تأخیر زمانی $h = |t - s|$ بستگی داشته باشد، سری زمانی ایستای ضعیف است که در این صورت تابع اتوکواریانس به $\gamma(h)$ خلاصه می شود. اندازه های دیگر و کلیدی عبارتند از توابع خود همبستگی و خود همبستگی جزئی در تأخیر h که برای سری های ایستا تعریف می شوند. تابع خود همبستگی میزان همبستگی خطی در تأخیر h و تابع خود همبستگی جزئی میزان همبستگی خطی در تأخیر h با حذف اثر زمان های بین را اندازه گیری می کنند. برآورد آنها نمودار های همبستگی نگار و همبستگی نگار جزئی را تشکیل می دهند که محور های اصلی در شناخت ویژگی یک سری ایستا می باشند. بر اساس رهیافت باکس-جنکینز سری های ایستا در کلاس مدل های فصلی ضربی اتورگرسیو- میانگین متحرک (ARMA) قابل مدل بندی هستند. مدل فصلی ضربی اتورگرسیو- میانگین متحرک مرتبه $(p, q) \times (P, Q)_s$ برای سری X_t رابطه خطی پیش بینی کننده ای به فرم زیر دارد:

$$\phi_p(B) \cdot \Phi_P(B) X_t = \theta_q(B) \cdot \Theta_Q(B) Z_t$$

که در آن Z_t بیانگر فرایند اغتشاش خالص و چند جمله ای های مشخصه $\phi_p(B)$, $\Phi_P(B)$, $\theta_q(B)$, $\Theta_Q(B)$ به ترتیب فصلی میانگین متحرک، غیر فصلی میانگین متحرک، فصلی اتورگرسیو، و غیر فصلی اتورگرسیو نامیده می شوند. همچنین عملگر B با تعریف $B X_t = X_{t-1}$ عملگر پس بر است. فرم کلی این چند جمله ای ها بصورت $1 \pm \sum_{i=1}^M \lambda_i B^i$ می باشد که در آن پارامتر S برای چند جمله ای های غیر فصلی برابر یک و برای چند جمله ای های فصلی برابر با طول دوره فصلی خواهد بود. پارامتر M متناظر با چند جمله ای مورد نظر، جانشین یکی از مرتبه های p, q, P, Q می باشد که تخمین آنها بوسیله رفتار همبستگی نگار و همبستگی نگار جزئی انجام می پذیرد. بعد از تخمین مرتبه ها، با برآورد ضرائب λ_i چند جمله ای های مشخصه و در نتیجه مدل برازشی معین می شود. با توجه به اینکه سری زمانی حاصل از پدیده های طبیعی مانند بارش دارای تغییراتی وابسته به زمان می باشند، انتظار می رود میانگین و واریانس آنها در طول زمان دچار تغییر شوند. این موضوع آنها را بطور مستقیم از گروه سری های ایستا خارج می کند. مدل های SARIMA بهترین پیشنهاد برای چنین سری هایی می باشند. سری غیر ایستای X_t در کلاس $SARIMA(p, d, q) \times (P, D, Q)_s$ قرار می گیرد اگر سری حاصل از تفاضل های فصلی و غیر فصلی آن که بصورت کلی $W_t = \nabla^d \nabla_S^D X_t$ قابل بیان

است یک فرایند ARMA فصلی از مرتبه $(P, Q)_S \times (p, q)$ باشد. عملگر تفاضلی ∇ و مرتبه n آن که به هدف حذف مؤلفه های فصلی و روند بکار گرفته می شود به صورت های $\nabla_S X_t = X_t - X_{t-S}$ و $\nabla_S^{n-1} X_t = \nabla_S^{n-1} (\nabla_S X_t)$ تعریف می شوند. ممکن است برای مشاهدات یک ایستگاه، معنی دار بودن چند الگوی آزمایشی به صورت همزمان پذیرفته شود که در این صورت برای انتخاب بهترین مدل از بین الگو های آزمایشی معیار اطلاعات آکائیک تصحیح شده (AICc) که فرم اصلاح شده معیار اطلاعات آکائیک معمولی از لحاظ اریبی می باشد، استفاده می شود [۵].

۲-۲- رهیافت شبکه های عصبی

شبکه های عصبی مصنوعی از چندین لایه به هم متصل از نرون ها تشکیل می شوند. نرون ها واحد های پردازشگری هستند که در یک لایه به صورت موازی عمل می کنند. هر شبکه از یک لایه ورودی و خروجی و احتمالا از چند لایه پنهان میانی تشکیل شده است. هر نرون وقتی بردار ورودی X را از نرون های قبلی دریافت می کند با یک عمل ماتریسی و تابعی آن را به نقطه y تبدیل کرده و خروجی را به نرون های بعد منتقل می کند. در این پردازش W و β بردارهای وزن از پیش تعیین شده و f تابع انتقال می باشد:

$$y = f(W.X - \beta) = f(\sum w_i x_i - \beta)$$

این فرآیند در لایه خروجی به برآیند نهایی رسیده و وزن های بهینه شبکه تعیین و تثبیت می شوند. چنین شبکه ای که در آن پیام ها تنها مسیر ورودی و خروجی به جلو را طی می کنند یک شبکه پیش رو خوانده می شود. روش تعیین مقادیر صحیح وزن ها را الگوریتم یادگیری می گویند. متداولترین الگوریتم یادگیری بنام پس انتشار خطا (BP°) نام گذاری شده است که در آن متناظر با ماتریس ورودی، ماتریس هدف نیز برای شبکه تعریف شده و فرآیند یادگیری تا زمانی ادامه می یابد که میانگین مربعات خطای شبکه به کمترین مقدار ممکن برسد. چنین شبکه ای علاوه بر لایه های ورودی خروجی، حداقل یک لایه پنهان نیز دارد. تعداد لایه ها، تعداد نرون های هر لایه، ماتریس وزن هر لایه، توابع انتقال و الگوریتم یادگیری عناصر تعیین کننده در طراحی یک شبکه عصبی می باشند.

^o -Back-Propagation

۲-۳ روش ترکیبی SARIMA و شبکه‌های عصبی

در روش ترکیبی ماتریس ورودی و هدف بر اساس تأخیرهای ظاهر شده در مدل های SARIMA، بعنوان بهترین پیش بینی کننده ها، انتخاب می شوند. تعداد نرون های لایه ورودی برابر با تعداد متغیرهای تأخیر است و لایه خروجی تنها دارای یک نرون با تابع انتقال خطی است. تعداد لایه های پنهان و نرون های آنها بصورت تجربی تا رسیدن به بهترین شبکه انتخاب می شوند. تابع انتقال آنها نیز از نوع تانژانت هایپربولیک یا لجستیک می باشد. لازم به توضیح است که معنی دار بودن ضرائب متغیرهای تأخیر بدست آمده در مدل SARIMA تنها در وجود یا عدم وجود آن تأخیر در ماتریس ورودی شبکه عصبی اهمیت دارد و مقدار عددی آنها تأثیری در نتایج طراحی شبکه عصبی ندارد.

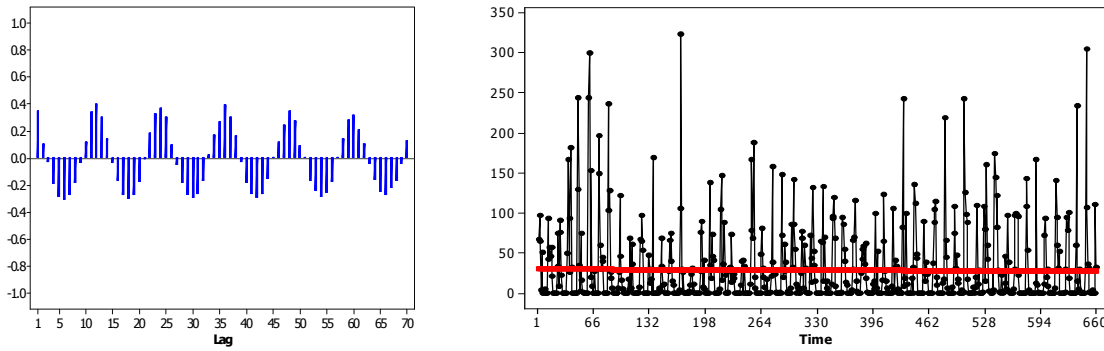
نتایج روش ترکیبی برای مشاهدات تحقیق

فرض کنید X_t میزان بارش ماهانه ایستگاه شیراز باشد. شکل (۱) نمودار سری زمانی مشاهدات X_t با خط روند و نمودار همبستگی نگار آن را نشان می دهد. همبستگی نگار وجود مؤلفه فصلی را تأیید می کند. اما مقدار احتمال بدست آمده در آزمون فرض صفر بودن شیب خط رگرسیون برای روند مساوی ۰/۷ بدست آمده است که بیان می کند وجود روند در سطح خطای ۰/۰۵ معنی دار نمی باشد. معادله خط روند به صورت زیر برآورد شده است:

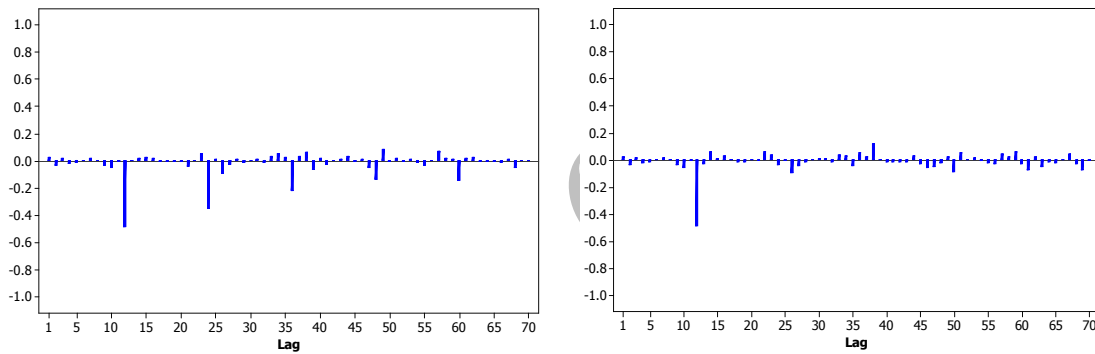
$$X(t) = 30.0 - 0.0004t, \quad R-Sq = 0.0\%$$

به وسیله عملگر تفاضلی با تأخیر ۱۲ مؤلفه فصلی حذف گردید. بنابر این مرتبه تفاضلی غیر فصلی (d) صفر و مرتبه تفاضلی فصلی (D) برابر یک در نظر گرفته شد. نمودار همبستگی نگار و همبستگی نگار جزئی سری های تفاضلی شده در شکل (۲) آمده است. همبستگی نگار تنها در تأخیر ۱۲ و همبستگی نگار جزئی در تأخیر های ۱۲ و ۲۴ و ۳۶ و ۴۸ و ۶۰ مقادیر معنی داری را نشان می دهند. بر این اساس، مرتبه های میانگین متحرک و اتورگرسیو غیر فصلی (p,q) صفر، مرتبه میانگین متحرک فصلی (Q) حداکثر یک و مرتبه اتورگرسیو فصلی (P) حداکثر پنج پیشنهاد می شود. بر اساس این تحلیل الگو های SARIMA با مرتبه های غیر فصلی و فصلی $Q=0,1$ و $P=0,1,2,3,4,5$ به مشاهدات برازش داده شده و از بین آنها مدل هایی با ضرائب معنی دار که فرض تصادفی محض بودن باقیمانده ها در آنها پذیرفته بود به عنوان الگو های

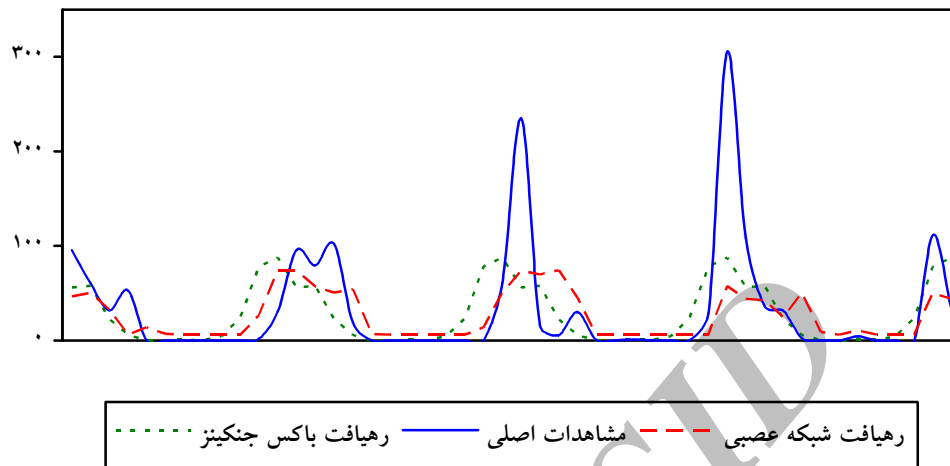
آزمایشی در نظر گرفته شدند. در نهایت مقایسه معیار آکاییک تصحیح شده در مدل های آزمایشی نشان از برتری مدل $SARIMA(0,0,0) \times (0,1,1)_{12}$ را داشت که با رابطه $\nabla_{12} X_t = (1 + \lambda B^{12}) Z_t$ بیان می شود. برآورد پارامتر λ در این مدل برابر $0/9236$ بدست آمد. نتیجه ای که از استخراج این مدل در طراحی شبکه عصبی اهمیت دارد این است که سری X_t تابعی از تأخیر X_{t-12} می باشد. بر همین اساس مشاهدات X_{t-12} به عنوان بردار ورودی و مشاهدات X_t به عنوان بردار هدف انتخاب شدند. به منظور آزمایش شبکه های طراحی شده و مقایسه نتایج پیش بینی آنها با رهیافت باکس-جنکینز، مشاهدات به دو قسمت تقسیم شدند. به گونه ای که ۵۰ سال اول که شامل ۶۰۰ مشاهده می باشد برای آموزش و ۵ سال آخر که شامل ۶۰ مشاهده می باشد برای آزمون شبکه ها استفاده گردید. بر این اساس مرتبه بردار ورودی و هدف در آموزش شبکه ها مساوی 1×588 و مرتبه بردار های آزمون شبکه ها مساوی 1×48 تعیین شد. پس از تعیین بردار های ورودی و هدف، شبکه های عصبی مختلفی با لایه ها و تعداد نرون های مختلف تا رسیدن به کمترین میانگین مربعات خطا آموزش داده شدند. پس از آموزش و تثبیت ماتریس های وزن، هر شبکه بوسیله بردار آزمون نیز مورد ارزیابی قرار گرفت. در شبکه های بهینه نتایج با پیش بینی های بدست آمده از رهیافت باکس-جنکینز برای ۴۸ مشاهده آخر مقایسه گردید و شبکه هایی بدست آمد که میانگین مربعات خطای آنها کمتر از رهیافت باکس-جنکینز بود. بهترین شبکه عصبی آموزش داده شده دارای یک لایه پنهان با ۶ نرون، تابع انتقال لجستیک برای لایه پنهان و خطی برای لایه خروجی بود. میانگین مربعات خطای پیش بینی این شبکه مساوی $2491/6$ بدست آمد که در مقایسه با مدل باکس-جنکینز که مساوی $2550/5$ می باشد مقدار کمتری را نشان می دهد. شکل (۳)، ۴۸ مشاهده آخر را بصورت توأم با پیش بینی های بدست آمده برای آنها از هر دو رهیافت شبکه های عصبی و باکس-جنکینز، نشان می دهد.



شکل ۱: نمودار سری زمانی X_t با خط روند (سمت راست) و همبستگی نگار آن (سمت چپ).



شکل ۲: نمودار همبستگی نگار (سمت راست) و همبستگی نگار جرئی (سمت چپ) سری تفاضلی شده $\nabla_1 X_t$.



شکل ۳: نمودار سری ۴۸ مشاهده اصلی آزمون در کنار پیش بینی های بدست آمده از هر دو رهیفات.

نتیجه گیری

ترکیب رهیفات باکس-جنکینز و شبکه های عصبی در مدل بندی سری های زمانی مورد ارزیابی قرار گرفت و به صورت موردی برای مشاهدات بارش ۵۵ سال ایستگاه سینوپتیک شیراز بکار گرفته شد. مدل مناسب از رهیفات باکس-جنکینز برای مشاهدات این ایستگاه بصورت $SARIMA(0,0,0) \times (0,1,1)_{12}$ ارائه شد. ماتریس های ورودی و هدف در طراحی شبکه های عصبی بر اساس تأخیر X_{t-12} تعیین و بوسیله آنها شبکه های عصبی مختلفی با پارامترهای مختلف تا رسیدن به کمترین میانگین مربع خطا آموزش داده شدند و نتایج بهتری نسبت به رهیفات باکس-جنکینز بدست آمد. بنابراین مدل ترکیبی برای پیش بینی بارش توصیه می گردد.

منابع

- ۱- Box G.E.P , Jenkins G.M. Time Series Analysis, Forecasting and Control, Revised edn.Holden-Day, San Francisco, CA, (۱۹۷۶).
- ۲- Buhamra S, Smaoui N, Gabr M, (۲۰۰۳). The Box-Jenkins analysis and Neural Networks: Prediction and Time Series Modeling. Applied Mathematical Modelling, ۲۷:۸۰۵- ۸۱۵.



- ۳- Faraway.J, Chatfield.C, (۱۹۹۸). Time Series Forecasting with Neural Networks: a Comparative Study using the air line data, Appl. statist, ۴۷:۲۳۱-۲۵۰.
- ۴- Sharifan, H., and Ghahraman, B. ۲۰۰۷. Evaluation of rainfall forecasting in Golestan province using time series. J. Agric. Sci. Natur. Resour. Vol. ۱۴, No. ۳: ۱۹۶-۲۰۹.(In Persian)
- ۵- Shumway R.H. and D.S. Stoffer. ۲۰۰۶. Time Series Analysis and its Applications with R Examples. Springer Science and Business Media, LLC.
- ۶- soltani, S., Modarres, R., and Eslamian, S.S. ۲۰۰۷. The use of time series modeling for determination of rainfall climates of Iran, Int. J. of Climatol. ۲۷: ۸۱۹-۸۲۹.

Archive of SID