

مدل سازی الگوی پخش فضایی ویروس کووید ۱۹ در ایران با شبکه عصبی پرسپترون چند لایه

لیلا مرویان مشهد، پرویز نصیری
گروه آمار، دانشگاه پیام نور تهران

چکیده: گاهی در مطالعات محیطی با مشاهداتی سرو کار داریم که مستقل از یکدیگر نیستند و نوعا وابستگی آنها ناشی از موقعیت و مکان قرار گرفتن مشاهدات در فضای مورد مطالعه می باشد. این گونه مشاهدات، داده های فضایی نامیده می شوند و به دلیل وجود همبستگی فضایی بین آنها، روش های معمول آمار برای تحلیل چنین داده هایی قابل استفاده نمی باشد و لازم است به نحوی ساختار همبستگی داده ها در تحلیل آنها لحاظ گردد. در مطالعات صورت گرفته در بیماری کووید-۱۹ مشاهده شده است که وضعیت شیوع این بیماری در نقاط مختلف بصورت داده های مستقل از یکدیگر نبوده و به مکان های مورد بررسی وابستگی دارد. مشاهدات بدست آمده بصورت داده های فضایی با همبستگی فضایی و عدم قطعیت می باشند که، روش های معمول آمار برای تحلیل چنین داده هایی از دقت کافی برخوردار نبوده و لازم است به نحوی ساختار همبستگی داده ها در تحلیل آنها لحاظ گردد. در این مقاله با استفاده از شبکه های عصبی پرسپترون چند لایه تلاش شده تا مدل بندی دقیق از شیوع بیماری کووید-۱۹ بدست آوریم. به این منظور از نرم افزار متلب جهت مدلسازی و شبیه سازی نتایج پیش بینی استفاده خواهیم نمود.

واژه های کلیدی: داده های فضایی، شبکه عصبی پرسپترون چند لایه، کووید ۱۹، همبستگی فضایی.
کد موضوع بندی ریاضی (۲۰۱۰): 60H11، 60G22.

۱ مقدمه

عدم قطعیت یکی از ویژگی های ذاتی در بسیاری از داده های زیستی، زمین آماری و جغرافیایی، به عنوان داده های فضایی است، که در اغلب اوقات ناشی از وجود خطا در اندازه گیری کمیت های مورد مطالعه است. این در حالی است که در نظر

نگرفتن این موضوع می‌تواند اعتبار نتایج حاصل از تحلیل داده‌ها را زیر سؤال برده و برآوردهای حاصل را دچار تورم واریانس و آریبی قابل توجهی نماید (اصغری، ۲۰۱۱).

شبکه‌های عصبی یکی از روش‌های ناپارامتری است که اخیراً در مدل‌سازی داده‌های فضایی به صورت گسترده‌ای مورد توجه قرار گرفته است. استفاده از این روش دارای مزایا و معایبی نیز است. از مزایای شبکه‌های عصبی می‌توان به عدم نیاز به فرضیات اولیه درباره مدل، قابلیت مدل‌سازی غیر خطی و انعطاف‌پذیری بالای آن اشاره کرد و از معایب شبکه‌های عصبی، نیاز به داده‌ای زیاد برای حصول نتایج دقیق‌تر و وابستگی شبکه به وزن‌های اولیه است. البته روشی برای تعیین تعداد داده‌ها، در دسترس نیست، اما معمولاً تعداد داده‌های مورد نیاز به ساختار شبکه و پیچیدگی مسئله مربوط می‌شود (الیوت و همکاران، ۱۹۹۶).

۲ مقدمه‌ای بر شبکه‌های عصبی

شبکه‌های عصبی مصنوعی (ANN) که با الگو گرفتن از دستگاه زیستی مغز انسان ساخته شده است، در ده‌های اخیر بسیار مورد توجه قرار گرفته است. مغز انسان که از تعداد بیشماری نورون تشکیل شده است با یک سیستم پردازش اطلاعات پیچیده، غیر خطی و موازی عملیات یادگیری پیچیده مانند طبقه‌بندی و شناسایی الگو را انجام می‌دهد. پیدایش شبکه‌های عصبی به اوایل اواخر قرن ۱۹ و قرن بیستم مربوط می‌شود. در آن دوره کارهای اساسی در فیزیک، روانشناسی و نرو فیزیولوژی توسط برخی دانشمندان چون ماخ و پاولف صورت پذیرفت که عموماً بر نظریه‌های کلی یادگیری تأکید داشتند و به مدل‌های ریاضی نورون اشاره‌ای نداشتند (فارو و دیگران، ۲۰۰۳). در دهه چهل قرن ۲۰ نشان داده شد که شبکه‌های عصبی می‌توانند هر تابع ریاضی و منطقی را تقریب بزنند. عملکرد این مدل مبتنی بر مجموع ورودی‌ها و ایجاد خروجی است به طوری که اگر حاصل جمع ورودی‌ها از آستانه‌ای معینی بیشتر باشد، نورون برانگیخته می‌شود. پس از آن مکانیسمی جهت یادگیری نورون‌های زیستی ارائه شد. در سال ۱۹۵۸ با معرفی شبکه پرسپترون نخستین کاربرد عملی شبکه‌های عصبی مطرح شد و شبکه‌ای ساخته شد که قادر بود الگوها را از هم شناسایی کند. در سال ۱۹۶۰ شبکه عصبی تطبیقی خطی یا قانون یادگیری جدید مطرح شد که از لحاظ ساختار شبیه شبکه پرسپترون است. در سال ۱۹۷۲ به طور مستقل شبکه‌های جدیدی معرفی شدند که به عنوان ذخیره ساز عمل می‌کردند. در این دهه روی شبکه‌های خودسازمانده فعالیت می‌شد. در دهه ۸۰ قرن بیستم به دلیل رشد فناوری ریزپردازنده‌ها، تحقیقات روی شبکه‌های عصبی افزایش یافت. در این دوره دو نظر جدید ارائه شد. اولین نظر که در سال ۱۹۸۲ مطرح شد، مکانیسمی تصادفی برای توضیح عملکرد بخش وسیعی از شبکه‌های برگشتی معرفی کرد که می‌توان از آن‌ها برای ذخیره سازی اطلاعات استفاده کرد. دومین نظر مهم، الگوریتم پس انتشار خطا بود. با بیان این دو نظر، شبکه‌های عصبی متحول شد. امروزه سرعت محاسباتی بالای رایانه‌ها و الگوریتم‌های یادگیری، باعث جذابیت شبکه‌های عصبی گردیده است.

۳ کاربرد شبکه عصبی در پیش‌بینی و مدل‌سازی داده‌های فضایی

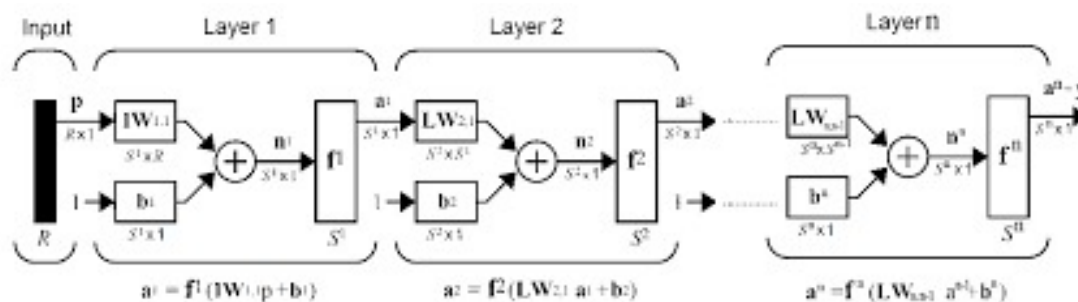
شبکه‌های عصبی مصنوعی یکی از روش‌هایی است که قادر به تخمین مدل‌های غیر خطی در داده‌ها هستند و اخیراً به طور گسترده‌ای در پیش‌بینی داده‌های فضایی مورد استفاده قرار گرفته‌اند. یکی از روش‌های متداول شبکه‌های عصبی در زمینه پیش‌بینی، شبکه عصبی پیش‌خور با یک لایه پنهان است. این نوع شبکه شامل ۳ لایه است: الف) لایه ورودی، ب) لایه پنهان و ج) لایه خروجی. ارتباط بین خروجی Y_t و ورودی‌های $(Y_{t-1}, Y_{t-2}, \dots, Y_{t-p})$ به صورت زیر بیان می‌شود:

$$Y_t = \alpha_0 + \sum_{j=1}^q \omega_{j,f} f(\beta_{0,j} + \sum_{i=1}^p \omega_{i,j} Y_{t-i}) + \varepsilon_t \quad (1.3)$$

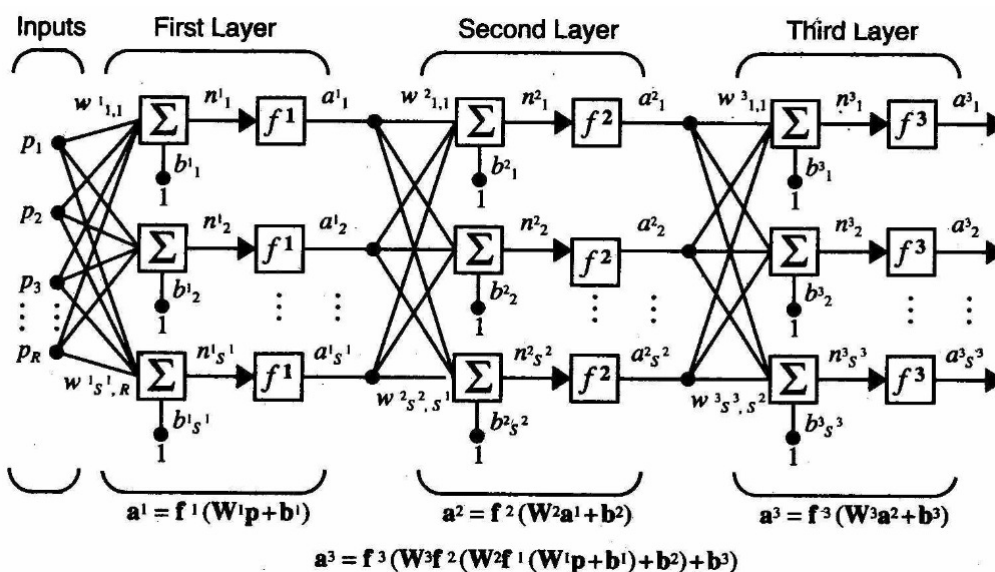
که در آن $w_{i,j}$ و w_j پارامترهای مدل هستند که اغلب به عنوان وزن‌های مدل شناخته می‌شوند و f تابع محرک سیگموئیدی است که اغلب به عنوان تابع انتقال در لایه‌های پنهان به کار می‌رود. در واقع روش شبکه‌های عصبی در رابطه (۱.۳) تابعی غیر خطی از مشاهدات گذشته $(Y_{t-1}, Y_{t-2}, \dots, Y_{t-p})$ به مشاهدات آینده Y_t توسط تابع زیر مربوط است:

$$Y_t = g((Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}, \omega) + \varepsilon_t$$

که در آن w وزن‌ها و g تابعی است که توسط ساختار شبکه و ارتباط وزن‌ها تعیین می‌شود. ساختار شبکه عصبی استفاده شده در این تحقیق به شکل زیر است. همانطور که مشاهده می‌شود این شبکه عصبی از سه لایه تشکیل شده است.



نحوه اتصال لایه‌ها و توابع وزنی این شبکه عصبی MLP را در زیر مشاهده می‌کنیم.



الگوریتم انتشار مجدد یکی از رایج‌ترین الگوریتم‌هایی است که از معماری چند لایه با تعاملات کامل بین لایه‌ها پیروی می‌کند. به طور معمول، یک یا چند لایه پنهان برای فعال کردن شبکه برای یادگیری وظایف پیچیده وجود دارد و آموزش بر اساس یک قانون یادگیری تصحیح خطا، یک قاعده کلی از الگوریتم کمترین میانگین مربع (LMS) است.

در مرحله آموزش، نورون‌های لایه پنهان و خروجی ورودی‌های خود را با ضرب هر ورودی در وزن مربوطه، جمع محصول، و سپس جمع‌آوری مجموع با استفاده از یک تابع انتقال غیر خطی برای تولید نتیجه پردازش می‌کنند. با تنظیم وزن بین نورون‌ها در پاسخ به خطاهای بین مقادیر خروجی واقعی و مقادیر خروجی هدف، یاد می‌گیرد. در پایان این مرحله آموزشی، شبکه عصبی طرحی را ارائه می‌دهد که باید بتواند یک مقدار هدف را از مقدار ورودی معین پیش‌بینی کند. به طور

کلی، روشی که برای برآورد مقادیر استفاده می‌شود. شبکه چند لایه را آموزش می‌دهد تا حداقل خطای هدفمند بین مقادیر خروجی مطلوب و واقعی به دست آید. پس از اتمام آموزش، مرحله پیش‌بینی دنبال می‌شود.

مرحله تحلیل فضایی:

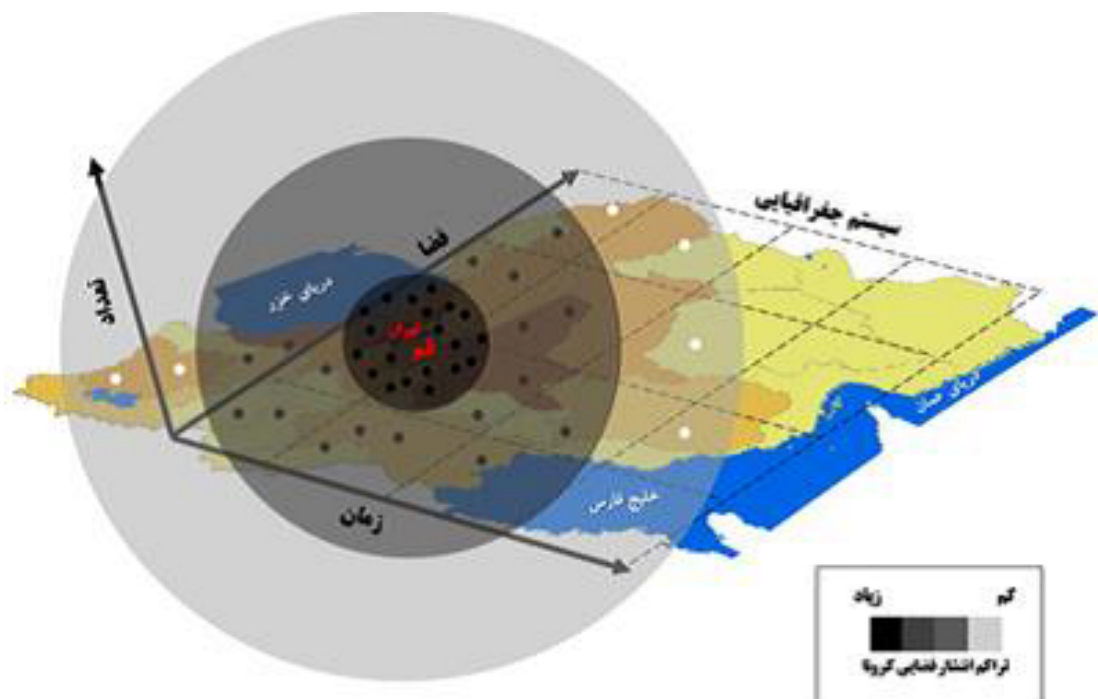
در مرحله پردازش داده‌ها، یک فایل متنی ساده شامل شماره شناسه، موقعیت توپوگرافی و میانگین مقادیر تعداد مبتلایان به Covid-19 هر استان مشاهده به منظور عادی‌سازی داده‌ها و ایجاد آموزش‌های لازم و مجموعه داده‌های اعتبارسنجی به سیستم وارد شد.

در مرحله آموزش، تمام عناصر پردازش لازم، تعداد سلول‌های عصبی، تعداد لایه‌های پنهان، میزان یادگیری و میزان حرکت، همه با مقادیر اولیه تعریف شده تنظیم شده است. این الگوریتم یک عملکرد بهینه را فعال می‌کند که به طور موازی داده‌های اعتبارسنجی را در مرحله آموزش نظارت می‌کند و با رسیدن به یک آستانه مشخص، فرآیند یادگیری را خاتمه می‌دهد. همچنین روشی برای مدیریت مجموعه داده‌های آموزشی موجود، تولید مجموعه‌ای از نتایج مختلف از طرح‌ها با عناصر پردازش متفاوت است. نتایج از طریق ماتریسی ارائه می‌شود که با کمک آن بهترین طرح پیش‌بینی را انتخاب می‌کنیم. هنگام عبور از مرحله آزمایش، پارامترهای شبکه عصبی سه لایه ذخیره می‌شوند (فرانکاویجیلیا، ۱۹۷۳). مجموعه‌ای آموزش داده نشده که به‌عنوان یک فایل متنی ساده نیز وارد می‌شود، عملکرد MLP را ارزیابی می‌کند.

در مرحله بعدی، مرحله پیش‌بینی، لایه‌ای که دارای اطلاعات مربوط به جنبه‌های توپوگرافی است، به شکل خاصی که مورد نیاز است به MLP معرفی می‌شود. پس از اجرای الگوریتم، نتایج دوباره به یک فایل شبکه تبدیل می‌شوند. مرحله نهایی، تحلیل فضایی، مرحله‌ای است که در آن فایل شبکه به یک بسته GIS وارد می‌شود و سطح پیوسته ایجاد شده از نظر بصری از نظر ناسازگاری یا خطا در حالی که پارامترهای آماری مشتق می‌شوند، بررسی می‌شود (بابلی و گاترل، ۱۹۹۵).

۴ بحث و نتیجه‌گیری

شکل زیر، مدلسازی الگوی فرآیند پخش فضایی کووید-۱۹ را در کشور نشان می‌دهد.



همانطور که مشاهده می‌شود، کانون اصلی این اپیدمی در کشور، استان قم است و از این استان به سایر استان‌های کشور، فرآیند پخش و انتشار فضایی ویروس کرونا صورت گرفته است. در نواحی پیرامون استان قم و به ویژه تهران (به دلیل وجود تراکم جمعیتی زیاد) شدت تراکم مبتلایان به این بیماری وجود دارد. با افزایش زمان و فاصله مکانی از کانون این بیماری (استان قم) از تعداد مبتلایان کاسته می‌شود. لذا به لحاظ ویژگی‌های جغرافیایی، دو عامل زمان و فاصله مکانی تأثیر زیادی در انتشار فضایی این ویروس در کشور داشته‌اند. به طوری که آمارها حاکی از آن است که استان بوشهر نسبت به سایر استان‌ها دیرتر درگیر این بیماری شده است.

بررسی‌های مکانی و جغرافیایی صورت گرفته نشان می‌دهد که مهمترین عوامل انتشار ویروس کرونا در کشور، تمرکز جمعیت و فاصله مکانی بوده که سبب سرعت بخشیدن به انتشار این بیماری واگیردار شده است. به عبارت دیگر، تا شعاع فاصله ۹۲۹ کیلومتری استان تهران انتشار ویروس کرونا روند صعودی دارد و از ۱۶۸ کیلومتری به دلیل افزایش فاصله مکانی و کاسته شدن از تراکم جمعیت انتشار بیماری کاهش می‌یابد. بررسی الگوی پراکنش انتشار ویروس کرونا در کشور نشان داد که الگوی فضایی این بیماری در کشور از نوع خوشه‌ای می‌باشد (کلیف، ۱۹۹۵). الگوی خوشه‌ای، بدترین الگوی پراکنش فضایی برای متغیرهایی مانند جرم و جنایت، بیماری، فقر، بیکاری و ... است که نشان دهنده بروز مشکل جدی در زمینه‌های مذکور بوده و نشان از حاد بودن آن مشکل در جامعه دارد. که در درجه اول باید با استفاده از همکاری مردم و سپس استفاده از توان علمی کشور به ویژه کادر پزشکی، نیروهای انتظامی و امنیتی، جغرافیدانان، روانشناسان و ... از ادامه روند کنونی این بیماری در کشور جلوگیری کرد.

مراجع

- Asgary, A., (2011), Space statistics analysis with GIS, First edition, Tehran: Municipality Information and Communication Technology Organization.
- Bailey, T., Gatrell, A., (1995), Interactive spatial data analysis, Harlow: Longman.
- Bell, B., Broemeling, L., (2000), A Bayesian analysis for spatial processes with application to disease mapping, Stat Med, Vol. 19, PP. 974-989.
- Cliff, A., (1995), Analyzing geographically related disease data, Stat Methods Med Res, Vol. 4, PP. 93-101.
- Dwyer, L., Burton, D., (1998), Potential meets reality: GIS & public health research in Australia, Aust J Public Health, Vol. 22, PP. 819-823.
- Elliott, P., Cuzik, J., English, D., Stern, R., (1996), Geographical & environmental epidemiology, 1st edition. England, Oxford University Press.
- ESRI, (2018), How Grouping Analysis works. <https://pro.arcgis.com/en/pro-app/toolreference/spatial-statistics/how-grouping-analysis-works.htm>. (26 May 2020).
- Faruque, F.S., Lofton, S.P., Doddato, T.M., Mangum, C., (2003), Utilizing Geographic Information systems in community assessment& nursing research, J Community Health Nurs, Vol. 20, PP. 179-191.

Francaviglia, R., (1973), Diffusion and popular Culture, An invitation to Geography, McGraw-Hill, London.

Ghaedamini Asadabadi, R., Tofighi, S., Ghaedamini, H., Azizian, F., Amerieon, A., Shokri, M., (2012), A review of some infectious diseases distribution based on geographic information system (GIS) in the area of Chahar Mahal and Bakhtiari, Journal of Police Medicine, Vol. 1, No. 2, PP. 113-123.