

پیش بینی بزرگی زلزله‌ی گسل تبریز با استفاده از رگرسیون چند جمله‌ای

علی خیری*: کارشناسی ارشد کامپیوتر - نرم افزار، پردیس بین المللی ارس دانشگاه تبریز، Email:aa.kh8905@gmail.com
محمد علی بالافر: استادیار، گروه مهندسی فناوری اطلاعات، دانشکده برق و کامپیوتر، دانشگاه تبریز.
بهزاد زمانی: استادیار، گروه علوم زمین، دانشکده علوم طبیعی، دانشگاه تبریز.

تاریخ دریافت: ۹۴/۶/۱۱

تاریخ پذیرش: ۹۵/۱۰/۲۵

چکیده

گسل تبریز یکی از گسل‌های فعال و خطرناک در ناحیه‌ی شمال غرب ایران است. این گسل از شمال شهر تبریز عبور کرده و شهرک‌های مختلفی را در معرض خطر قرار داده است؛ همچنین بخش اعظمی از ساکنان حاشیه‌نشین شهر تبریز در قسمتی از کمربند شمالی شهر اسکان یافته‌اند که ساختمان‌های آن‌ها به هیچ‌وجه در مقابل کوچک‌ترین تکان‌ها و زمین‌لرزه‌ها مقاوم نیست. بررسی داده‌های لرزه‌ای گسل تبریز با استفاده از رگرسیون چند جمله‌ای می‌تواند با برآورد میزان خطرات احتمالی و بزرگی زلزله‌های پیش رو برای پیشگیری از خطرات ناشی از زلزله مفید و سودمند باشد. در گام نخست همه‌ی داده‌های ورودی استخراج و نرمالیزه شده و در گام دوم مدل مورد نظر به دست می‌آید. در گام سوم بزرگی زلزله‌ها برای داده‌های تست محاسبه می‌گردد و در گام چهارم کیفیت پیش‌بینی بزرگی زمین‌لرزه‌ها بررسی می‌شود. در نهایت برای تخمین مقدار بزرگی زلزله‌ها با کمک دیگر صفات، رابطه‌ای به دست آمده است که می‌تواند ضرایب هر یک از متغیرهای مستقل برآوردکننده‌ی بزرگی زلزله را بیان نماید.

واژه‌های کلیدی: داده‌کاو، رگرسیون چند جمله‌ای، داده‌های لرزه‌ای، گسل، پیش‌بینی

Prediction of Tabriz fault's earthquakes magnitude using polynomial regression

Ali kheiri^{*1}, Mohamad ali Balafar², behzad Zamani³

Abstract

Tabriz fault is one of the active and dangerous faults in the North West of Iran. The fault crosses the north through the city and put crossing towns at risk. The majority of marginal inhabitants within the city are settled there. Their buildings do not resist the slightest shake. Analysis of seismic data using polynomial regression for predicting future earthquake's magnitude and risks can be helpful to reduce disaster results. First of all, input data are extracted and normalized. At the second step, model is obtained. In the third step, the magnitude of earthquake for the test data is predicted. In the fourth step, the performance of predicting earthquake magnitude is evaluated. Finally, to estimate the magnitude of earthquakes using other attributes a formula is obtained that can be used to estimate the coefficients of each of the independent variables.

Keywords: data mining, polynomial regression, seismic data, fault, prediction.

- 1 Graduated master degree Computer software, Aras International Campus, University of Tabriz, Tabriz, Iran; Email:aa.kh8905@gmail.com
- 2 Assistant Professor, Faculty of electrical & computer engineering, University of Tabriz, Tabriz, Iran.
- 3 Assistant Professor Faculty of Natural sciences University of Tabriz, Tabriz, Iran.

۷۷

شماره دهم

پاییز و زمستان
۱۳۹۵

دوفصلنامه
علمی و پژوهشی



ه. ش (۱۷۸۰ م.) به منزله‌ی یکی از مخرب‌ترین و مرگ‌بارترین زلزله‌های جهان با ۷۷ هزار کشته [۱] خودنمایی می‌کند. در تصویر ۱ گسل تبریز به همراه محل رخدادهای آن آورده شده است.

بنابراین تحلیل و بررسی داده‌های زمین‌لرزه‌ای گسل تبریز از اهمیت ویژه‌ای برخوردار است. بحث حرکت‌شناسی زلزله و مدل‌سازی رایانه‌ای آن، یکی از چالش‌برانگیزترین مسائل در ژئوفیزیک مدرن است. هرچند زلزله‌شناسان پس از وقوع زلزله‌های بزرگ قادر به ارائه‌ی پیش‌گویی‌های موفقیت‌آمیزی بوده‌اند، اما هنوز هم پیش‌بینی زلزله به علت پیچیدگی بالا، کار سختی به شمار می‌آید [۲]. با استفاده از مدل‌سازی و شبیه‌سازی رایانه‌ای و روش‌های داده‌کاوی^۳ می‌توان این پدیده‌ها را مطالعه نمود و از روش‌های یادگیری خودکار برای استخراج ارتباط زلزله‌ها با پدیده‌های مختلف استفاده کرد تا پیش‌بینی‌های درست‌تری به دست آید. کشورهایی مثل چین و ژاپن که از مناطق زلزله‌خیز جهان به حساب می‌آیند به پیش‌بینی و مقاوم‌سازی در برابر زلزله توجه زیادی داشته‌اند [۳]. از آنجا که کشور ما و به‌ویژه شهر تبریز از مناطق با لرزه‌خیزی بالا محسوب می‌شود، این موضوع باید با جدیت بیشتری پی‌گیری شود.

روش تحقیق و ابزارها

علم داده‌کاوی از علوم مختلفی از جمله آمار، هوش مصنوعی، یادگیری ماشین، شناسایی الگو و پایگاه داده نشأت گرفته است. در واقع این علوم ریشه‌های علم داده‌کاوی هستند. ما می‌توانیم از همه‌ی روش‌های موجود در این علم بهره بگیریم به گونه‌ای که بتوانیم بفهمیم الگوریتم‌ها و روش‌هایی که در داده‌کاوی هستند چگونه عمل می‌کنند. الگوریتم‌های موجود در هوش مصنوعی و علم آمار کمک شایانی به داده‌کاوی می‌کنند. داده‌کاوی عبارت است از فرایند خودکار کشف دانش و اطلاعات از پایگاه‌های داده‌ای. از مشهورترین روش‌های یادگیری مدل در داده‌کاوی می‌توان به روش‌های پیش‌بینی و روش‌های توصیفی اشاره کرد. روش‌های پیش‌بینی از مقادیر بعضی از ویژگی‌ها برای پیش‌بینی مقدار یک ویژگی مشخص استفاده می‌کنند. در متون علمی مختلف روش‌های پیش‌بینی با نام روش‌های ناظر نیز شناخته

مقدمه

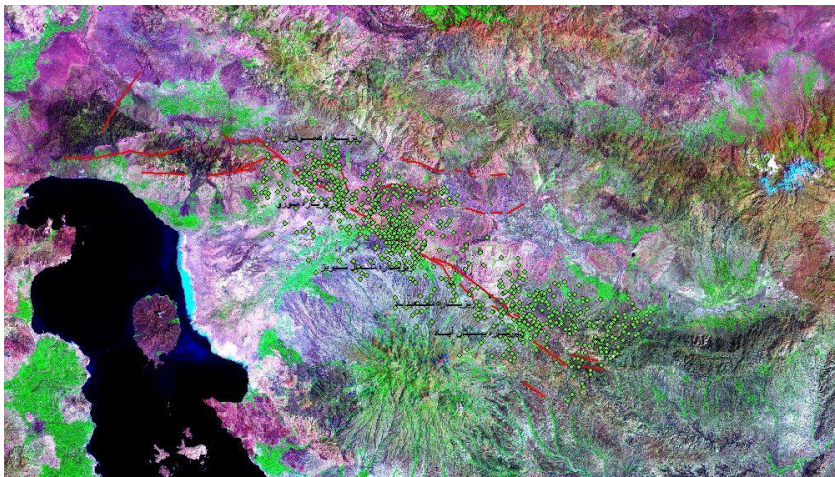
منطقه‌ی شمال غرب ایران، به دلیل عملکرد گسل‌های بزرگ و کوچک با روندهای متفاوت که بزرگ‌ترین آن‌ها گسل شمال تبریز است، دارای ساختار تکتونیکی^۱ پیچیده‌ای است که به منزله‌ی قسمتی از کمربند آلپ-همالیا دارای گسل‌های فعال و زمین‌لرزه‌های تاریخی ویرانگر بوده و از نواحی پویا و لرزه‌خیز فلات ایران به حساب می‌آید. گسل شمال تبریز، برجسته‌ترین و مهم‌ترین سیمای ساختاری منطقه به‌منزله‌ی یک پهنه با حرکات غالب امتدادلغز راست بر با روند شمال غرب-جنوب شرق، اساسی‌ترین نقش را در تکوین و تکامل ساختاری، دگرشکلی و خواص مورفولوژیکی^۲ منطقه، طی دوره‌ی بلندمدت زمین‌شناسی داشته است. این گسل با پیشینه‌ی لرزه‌خیزی طولانی در شمال غرب کشور، به لحاظ قرارگیری شهر تبریز به‌منزله‌ی یکی از شهرهای پرجمعیت کشور در کنار آن همواره مورد توجه بوده است. گسل تبریز از شمال این شهر عبور کرده و شهرک‌های مختلفی را در معرض خطر قرار داده است، همچنین بخش اعظمی از حدود هزاران حاشیه‌نشین شهرستان تبریز در قسمتی از کمربند شمالی شهر اسکان یافته‌اند که ساختمان‌های آن‌ها به‌هیچ‌وجه در مقابل کوچک‌ترین تکان‌ها نیز مقاوم نیست. این سامانه گسلی به طول تقریبی ۱۵۰ کیلومتر از حوالی مرند در شمال غرب تا بستان‌آباد در جنوب شرق ادامه دارد و علت اصلی وقوع زمین‌لرزه‌های تبریز در گذشته، حرکت و جنبش این گسل بوده است. گسل تبریز به علت بیشترین ساخت و ساز و تراکم جمعیت روی آن یکی از خطرناک‌ترین گسل‌های ایران است و در صورت رخ دادن زلزله در تبریز، شهر با تکان‌های سهمگین مواجه می‌شود ولی شمال و شرق این کلانشهر که روی گسل واقع شده است، دچار گسیختگی‌های ویرانگر زمین و زمین‌لغزش‌های متعدد خواهد شد. با وجود گسل شمال تبریز وقوع زمین‌لرزه‌های بزرگ در آینده‌ای نامعلوم واقعیتی انکارناپذیر است. در زمان گسیختگی زمین، هیچ بنای انسان‌ساز مقاومت ندارد و تخریب می‌شود. این شهر بارها در طی تاریخ در اثر زلزله ویران شده است. برای واقف شدن به اهمیت موضوع زمین‌لرزه‌های شهر تبریز، ذکر این نکته کافی است که زلزله‌ی ۱۸دی ماه سال ۱۱۵۸

۷۸

شماره دهم
پاییز و زمستان
۱۳۹۵

دوفصلنامه
علمی و پژوهشی

پیش‌بینی
زلزله‌ها



تصویر ۱: گسل تبریز (محل رخدادهای زمین‌لرزه‌ها)

چندجمله‌ای
پیش‌بینی بزرگی زلزله‌ی گسل تبریز با استفاده از رگرسیون

می‌شوند [۸]. روش‌های دسته‌بندی^۴، رگرسیون^۵ و تشخیص انحراف^۶ سه روش یادگیری مدل در داده‌کاوی با ماهیت پیش‌بینی هستند. روش‌های توصیفی الگوهای قابل توصیفی را پیدا می‌کنند که روابط حاکم بر داده‌ها را بدون در نظر گرفتن هرگونه برچسب و یا متغیر خروجی تبیین نمایند. در متون علمی مختلف روش‌های توصیفی با نام روش‌های بدون ناظر نیز شناخته می‌شوند [۸]. روش‌های خوشه‌بندی^۷، کاوش قوانین انجمنی^۸ و کشف الگوهای تزیینی^۹ سه روش یادگیری مدل در داده‌کاوی با ماهیت توصیفی هستند.

رگرسیون چندجمله‌ای یکی از قدرتمندترین روش‌های آماری فرایند داده‌کاوی است. رگرسیون خطی داده‌ها را به صورت یک خط راست مدل‌سازی می‌کند. خط رگرسیون ابزاری است برای پیش‌بینی مقدار یک متغیر بر حسب متغیری که به آن وابسته است. در واقع برای مدل‌سازی مقادیر دو صفت خاصه، خطی را می‌یابیم که به هم‌ی زوج مقادیر این دو صفت خاصه نزدیک باشد. چنانچه زوج‌های مرتب روی یک خط راست نباشند، برای هر زوج و خط رگرسیون خطایی وجود دارد. این خط به نحوی انتخاب خواهد شد تا خطا به حداقل خود برسد. رگرسیون چندگانه حالت توسعه‌یافته‌ای از رگرسیون خطی است، با این تفاوت که در آن بیش از دو متغیر وجود دارد و مقدار یکی با توجه به دیگر متغیرها تخمین زده می‌شود. به حداقل رساندن ریشه‌ی مجموع مربعات خطاها روش مرسوم است که در بیشتر موارد از آن استفاده می‌شود.

تحلیل رگرسیونی روشی است که برای مطالعه‌ی روابط بین متغیرها و به‌ویژه درک چگونگی وابستگی یک متغیر به سایر متغیرها کاربرد دارد. به عبارتی تحلیل رگرسیونی تحلیلی برای کمی نمودن ارتباط بین یک متغیر هدف (وابسته) و یک یا چند متغیر پیش‌بینی‌کننده (مستقل) است.

پیشینه‌ی پژوهش

برای مقابله با تهدیدات لرزه‌ای روش‌ها و ابزارهای مختلف برای پیش‌بینی مکان، زمان و بزرگای زلزله به کار گرفته شده است. طراحی یک سیستم پشتیبانی تصمیم‌گیری^{۱۰} هدف اصلی این مقاله است. این سیستم با بهره‌برداری از روش‌های داده‌کاوی و با استفاده از یک مدل به پیش‌بینی می‌پردازد و از سه بخش اصلی تشکیل یافته است: فرایند مدیریت بارگذاری که از SQL سرور برای سرور پایگاه داده و از وکا^{۱۱} برای داده‌کاوی استفاده می‌کند و شمای چندبعدی که از مدل‌سازی چندبعدی پایین به بالا استفاده می‌کند و در نهایت عملیات OLAP را انجام می‌دهد. این تکنولوژی داده‌های موجود را به صورت چندبعدی تجزیه و تحلیل می‌کند و بدین ترتیب توانایی انجام محاسبات پیچیده، تجزیه و تحلیل رویه‌ها و مدل‌سازی داده‌های پیچیده را فراهم می‌کند. مدل‌های توصیفی و پیش‌بینی که در بخش‌های بالا توصیف شده است، از روش k-means برای خوشه‌بندی داده‌ها و از الگوریتم آپریوری^{۱۲} برای قوانین معنایی استفاده می‌کند تا شمای مدل توصیفی را کامل نماید. در مقابل مدل پیش‌بینی نیز، از

درخت‌های کلاس‌بندی شده و رگرسیون^{۱۳} استفاده می‌کند. فرض بر این است که اگر ناحیه‌ای در بازه‌های زمانی کوتاه، چندین زلزله را تجربه کند، این ناحیه مستعد رویدادهای بزرگ نیز هست. فهرست داده‌ای اصلی استفاده شده در مقاله چنین است:

US Geological Survey (USGS), National Earthquake Information (NEIC), National Siesmic System of Mexico (SSN) (Smodevilla, 2010)

میانگین بزرگای رویدادهای مطالعه شده، ۶/۱۶ است. نمودارهای به دست آمده بر اساس ساعت و روز در نظر گرفته شده است که بیشترین تمرکز زلزله در شانزدهمین روز از ماه و معمولاً قبل از ۴ صبح است. البته قابل ذکر است که این نظریه به صورت علمی ثابت شده نیست و فقط توسط مشاهدات آماری به دست آمده است [۴].

تئوری و محاسبات

داده‌های لرزه‌ای که در این پژوهش مورد استفاده قرار گرفته است شامل فعالیت‌های لرزه‌ای و مشخصات زلزله‌های دستگاهی است. تاکنون کاتالوگی که تمام زمین‌لرزه‌های سده‌ی بیستم را در برگیرد انتشار نیافته است و در حال حاضر داده‌های دهه‌ی آغازین این سده در دست نیست. این داده‌های لرزه‌ای ترکیبی است از داده‌های بولتن زلزله‌شناسی مؤسسه‌ی ژئوفیزیک دانشگاه تهران (IRSC) [۵] و بولتن زلزله‌شناسی و کاتالوگ زمین‌لرزه‌های ایران که توسط شبکه‌ی ملی لرزه‌نگاری باند پهن ایران در پژوهشگاه بین‌المللی زلزله‌شناسی و مهندسی زلزله (IIIES) [۶]، تهیه شده است. این کاتالوگ مشخصات کامل زمین‌لرزه‌های منتسب به گسل شمال تبریز را از سال ۱۹۱۴ میلادی تا ۲۰۱۳ میلادی شامل می‌شود. این داده‌ها شامل ویژگی‌های زمان وقوع زمین‌لرزه، بزرگای زمین‌لرزه بر اساس مقیاس‌های امواج درونی (Mb)، امواج سطحی (Ms)، گشتاوری (Mw) و مقیاس محلی (ML) و همچنین عمق کانونی، طول و عرض جغرافیایی، شماره‌ی ایستگاه و سایر ویژگی‌های زمین‌لرزه هستند. از بین عوامل متعدد جمع‌آوری شده برای این داده‌ها از نظر زمین‌شناسی فقط پارامترهای زمان، طول و عرض جغرافیایی، عمق و بزرگی رویدادها به منزله‌ی مؤلفه‌های مفید در پیش‌بینی زلزله انتخاب شدند که مقادیر ورودی و محدوده‌ی هر کدام در جدول ۱ آمده است.

جدول ۱: محدوده‌ی پارامترهای استفاده شده در این تحقیق

داده‌های مورد استفاده	کمترین	بیشترین
زمان رویداد (Date)	۱۹۱۴	۲۰۱۳
بزرگای زلزله (M _b)	۲/۳	۵/۸
عمق کانونی (Depths)	۰	۳۵
طول جغرافیایی (Long)	۴۵/۱	۴۷/۳
عرض جغرافیایی (Lat)	۳۷/۷۱	۳۸/۷۲

با توجه به داده‌های جمع‌آوری شده پس از اعمال پیش‌پردازش بر روی این داده‌ها و حذف پیش‌لرزه‌ها و پس‌لرزه‌ها از لرزه‌های اصلی، انتخاب آن‌ها بر اساس کامل بودن پارامترهای

مورد نیاز، برای داده‌های دستگامی، در نهایت از بین حدود ۱۳۴۷ زمین‌لرزه مربوط به سال‌های ۱۹۱۴ تا ۲۰۱۳ میلادی، ۱۳۸ زمین‌لرزه‌ی اصلی به دست آمد.

در مورد زمان وقوع رویدادها که شامل روز و ماه و سال است، همگی تبدیل به روز شده و از تاریخ ۱۹۰۰/۱/۱ به عنوان تاریخ مبدأ در نظر گرفته شده و به فرمت عددی در آمده‌اند. به این ترتیب مقدار حاصل، علاوه بر این که ترتیب رویدادها را نشان می‌دهد میزان فاصله‌ی زمانی هر دو رویداد متفاوت از همدیگر را نیز حفظ می‌نماید. همچنین بزرگای رویدادها همه به مقیاسی یکسان که همان M_w (گشتاوری) است درآمدند و روش یکسان‌سازی بزرگاها با استفاده از مقاله‌ی بولت به دست آمدند [۹]. گفتنی است که در این مقاله همواره منظور از برآورد بزرگا همان برآورد M_w است. پس از جمع‌آوری داده‌های لرزه‌ای گسل تبریز با استفاده از نرم‌افزار داده‌کاوی Rapid Miner و بر اساس الگوریتم رگرسیون چندجمله‌ای^{۱۴} اقدام به پیش‌بینی بزرگی زمین‌لرزه‌های آتی گسل تبریز نمودیم. برای مدل‌سازی بر اساس رگرسیون چندجمله‌ای برای برآورد بزرگای زمین‌لرزه‌ها، از داده‌هایی شامل توالی زمانی وقوع زمین‌لرزه‌ها، طول و عرض جغرافیایی، عمق کانونی و بزرگای زمین‌لرزه‌ها بهره گرفته شده است. مجموع داده‌های استفاده شده در این بخش، برای پیش‌بینی بزرگا شامل ۱۳۸ داده‌ی لرزه‌ای گسل تبریز است. الگوریتم رگرسیون چندجمله‌ای را بر روی داده‌های لرزه‌ای اعمال کردیم. برای ارزیابی درستی مدل به دست آمده از دو معیار ارزیابی به نام‌های Split-Validation و X-Validation استفاده کردیم که سازوکار کار هر کدام در بخش بعدی آمده است. در جدول ۲ نمونه‌ای از داده‌های مورد استفاده آورده شده است.

بحث و نتایج

روش کار Split Validation به این صورت است که در این روش برای تقسیم مجموع داده‌ها به دو بخش آموزشی و آزمایشی می‌توان سهم هر بخش را به صورت دلخواه و با هر نسبتی تعیین کرد [۷] که ما در این تحقیق از ۷۰ درصد داده‌ها برای آموزش و ۳۰ درصد داده‌ها برای آزمایش استفاده نمودیم.

جدول ۲: نمونه‌ای از داده‌های استفاده شده

Date	Lat	Long	Depth	M_w
۲۰۰۷/۰۹/۱۶	۳۸/۱۰۴	۴۶/۳۷۷	۱۵/۶	۴/۴
۲۰۰۷/۱۲/۰۱	۳۸/۱۷۲	۴۶/۳۹۳	۴/۲	۴/۶
۲۰۰۸/۰۱/۲۲	۳۸/۰۷۳	۴۶/۳۳۷	۱۰	۲/۵
۲۰۰۸/۰۱/۲۸	۳۷/۸۱۶	۴۶/۹۷۲	۵/۴	۲/۸
۲۰۰۸/۰۲/۲۳	۳۷/۸۲۶	۴۷/۰۷۷	۵/۵	۲/۷
۲۰۱۱/۰۴/۰۱	۳۷/۸۳	۴۶/۹۹۸	۸/۹	۳/۳
۲۰۱۱/۰۵/۰۴	۳۸/۲۹	۴۵/۹۵۹	۵/۳	۲/۳
۲۰۱۲/۰۹/۱۸	۳۷/۸۴۸	۴۶/۶۵۷	۵/۸	۲/۹
۲۰۱۲/۱۲/۱۶	۳۷/۹۲۹	۴۶/۷۵۲	۱۰	۳/۲
۲۰۱۳/۰۲/۰۸	۳۷/۸	۴۷/۰۲۸	۱۰	۲/۶

اما در روش X-Validation مجموع داده به x قسمت برابر تقسیم می‌شود، با $x-1$ قسمت عمل آموزش و با یک قسمت باقی‌مانده عمل ارزیابی انجام خواهد شد. این کار تا تمام شدن تمام حالات موجود یعنی x بار تکرار می‌شود، به گونه‌ای که از هر کدام از x قسمت تنها یک بار برای ارزیابی استفاده شود و در هر مرتبه یک دقت برای مدل ساخته شده محاسبه شود. در این روش ارزیابی دقت نهایی پیش‌بینی برابر با میانگین x دقت محاسبه شده خواهد بود [۸].

نتایج به دست آمده پس از اعمال الگوریتم رگرسیون چندجمله‌ای بر روی داده‌های لرزه‌ای گسل تبریز با دو روش ارزیابی Split Validation و X-Validation با ترکیب متفاوتی از صفت‌های مستقل برآوردکننده و بعد از نرمال‌سازی داده‌ها در بازه‌ی صفر تا یک در جدول ۳ آمده است.

یکی از مهم‌ترین عواملی که در تجمع تنش در فاصله‌ی دو رویداد لرزه‌ای (در فاصله‌ی دوره‌ی بازگشت زلزله‌ها) وجود دارد تجمع تنش‌ها است. در واقع در فاصله‌ی زمانی دوره‌ی بازگشت یک زلزله‌ی مشخص بر مبنای مدل لرزه‌خیزی Stick-slip، تنش‌ها تجمع یافته و به مرحله‌ی گسیختگی مجدد در یک پاره گسل مشخص می‌گردند. از آنجا که در بیشتر موارد به علت کمبود و یا نبود داده‌های لازم برای محاسبه‌ی بزرگی تنش‌های حاکم بر یک پاره از گسل امکان محاسبه و مقایسه‌ی تنش‌ها برای هر رویداد زلزله وجود ندارد، بنابراین حداقل می‌شود از عامل زمان رویداد زلزله‌ها که یکی از دقیق‌ترین داده‌های لرزه‌ای در ایران محسوب می‌شوند، بهره جست. در این ارتباط از عامل زمان رویداد در کنار بزرگی زلزله‌ها در تحلیل رگرسیون چندجمله‌ای به‌منزله‌ی عامل مؤثر در تجمع تنش بهره گرفته شده است [۱۰].

جدول ۳: مقادیر به دست آمده از رگرسیون چندجمله‌ای با معیار ارزیابی Split-Validation و X-Validation

مدل	ریشه‌ی مجموع مربعات خطاها		صفات مستقل برآورد کننده
	X-Validation	Split-Validation	
۱	۰/۱۶۰	۰/۲۳۶	طول و عرض جغرافیایی، زمان
۲	۰/۲۲۳	۰/۲۲۸	طول و عرض جغرافیایی، عمق
۳	۰/۱۷۱	۰/۱۶۴	زمان، عمق
۴	۰/۱۶۰	۰/۱۵۲	زمان، عمق، طول و عرض جغرافیایی

برای پیش‌بینی بزرگی زمین‌لرزه‌ها با استفاده از روش رگرسیون چندجمله‌ای از روی صفات مستقل برآوردکننده‌ی بزرگی مطابق جدول ۱ در روش ارزیابی X-Validation، با مقدار $x=10$ که در بسیاری از کاربردها تعیین مقدار ۱۰ توصیه شده است [۷] و در روش ارزیابی Split Validation از ۷۰ درصد داده‌ها برای آموزش و ۳۰ درصد باقی‌مانده‌ی داده‌ها برای آزمایش مدل با نحوه‌ی نمونه‌برداری برهم‌آمیخته^{۱۵} و تصادفی استفاده کردیم. چنان‌که مشاهده می‌شود بهترین حالت پیش‌بینی با کمترین ریشه‌ی

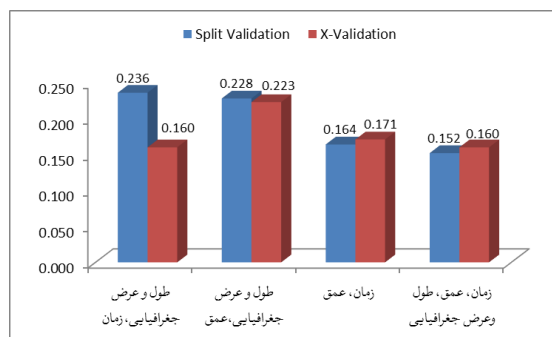
پی‌نوشت

1. Tectonic
2. Morphological
3. Data Mining
4. Classification
5. Regression
6. Anomaly Detection
7. Clustering
8. Association Rule Mining
9. Sequential Pattern Discovery
10. Decision Support System
11. Weka
12. Apriori
13. Regression
14. Polynomial Regression
15. shuffled

منابع

1. بهیاری، مهدی؛ محجل، محمد مهدی (۱۳۸۶). نرخ لغزش کلی گسل تبریز و رابطه‌ی آن با تئوری تکتونیک فراری بر شمال غرب ایران. بیست و ششمین گردهمایی علوم زمین، تهران، وزارت صنایع و معادن، سازمان زمین‌شناسی و اکتشافات معدنی کشور.
2. Bose, M. Wenzel, F (2008). Preseis: A Neural Network-Based Approach to Earthquake Early Warning for Finit Faults. *Bulletin of the Seismological Society of America*, 98 (1), 366-382.
3. Shengkuzi, Z. Chengmin, W (1999). Application of Artificial Intelligence in Earthquake forecasting. In 1st ACES Workshop proceedings, 477-481.
4. Smodevilla, M.j. Priego, A.B (2010). Decision Support System for Seismic Risks. *Journal of Computer Science & Technology*, (JCS&T), 12 (2), 71-77.
5. <http://irsc.ut.sc.ir/index.php?lang=fa> (2013).
6. www.iiees.ac.ir (2013).
7. اسماعیلی، مهدی (۱۳۹۲). آموزش گام به گام داده‌کاوی با Rapid Miner. تهران، آتی‌نگر.
8. صنیعی آباده، محمد؛ محمودی، سینا؛ طاهرپرور، محدثه (۱۳۹۳). داده‌کاوی کاربردی (ویرایش دوم). تهران، نیاز دانش.
9. Bolt B.A. (1993). *Earthquakes*. New York, W.H. Freeman and Company.
10. Shcolz. (1987). *Earthquakes and friction laws*. Nature 391, 37-42.

مجموع مربعات خطا در معیار ارزیابی Split Validation با استفاده از متغیرهای مستقل برآوردکننده‌ی زمان، عمق، طول و عرض جغرافیایی با مقدار ۰/۱۵۲ - اتفاق افتاده است. برای درک بهتر مدل، نتایج حاصل در تصویر ۲ به صورت نموداری آمده است.



تصویر ۲: مقایسه‌ی مقادیر کمترین ریشه‌ی مجموع مربعات خطا به دست آمده از رگرسیون چند جمله‌ای با معیار ارزیابی Split-Validation و Validation

رگرسیون چندجمله‌ای برای پیش‌بینی بزرگی زمین‌لرزه‌های گسل تبریز برای هر مدل جدول ۱ رابطه‌ای را ارائه می‌دهد که دو نمونه از رابطه‌های به دست آمده برای حالت‌های بهینه‌ی روش‌های اعتبارسنجی Split-Validation و X-Validation آورده می‌شود. معادله‌ی بهینه‌ی به دست آمده از معیار ارزیابی Split Validation برابر است با:

$$Mw = (-0.905 * Date^1) + (0.351 * Lat^1) + (0.178 * Long^3) + (0.157 * Depth^6) + 0.858$$

و همچنین رابطه‌ی بهینه‌ی به دست آمده از روش ارزیابی X-Validation با صفت‌های مستقل زمان وقوع و طول و عرض جغرافیایی رویدادها برابر است با:

$$Mw = (-0.797 * Date^2) + (0.178 * Lat^2) + (0.040 * Long^1) + 0.843$$

با استفاده از رابطه‌های بالا می‌توان بزرگی زمین‌لرزه‌ها را برای داده‌های تست به دست آورد.

نتیجه‌گیری

برای به دست آوردن بزرگی زمین‌لرزه‌های آتی با استفاده از ۱۳۸ داده‌ی لرزه‌ای گسل تبریز و الگوریتم رگرسیون چندجمله‌ای بهترین حالت پیش‌بینی مربوط به معیار ارزیابی Split Validation با صفت‌های مستقل برآوردکننده‌ی بزرگی نظیر زمان، عمق، طول و عرض جغرافیایی رویدادها با کمترین ریشه‌ی مجموع مربعات خطا برابر با ۰/۱۵۲ شد. با توجه به رابطه‌ی به دست آمده برای این مدل، می‌توان نتیجه گرفت که زمان وقوع زمین‌لرزه‌ها مهم‌ترین متغیر مستقل برآوردکننده‌ی بزرگی است. با استفاده از رابطه‌ی ۱ که رابطه‌ی بهینه‌ی بین تمام حالات است با وارد کردن داده‌های لرزه‌ای جدید می‌توان بزرگی زمین‌لرزه را با کمترین خطا پیش‌بینی کرد.