

ارائه روشی مبتنی بر برنامه نویسی ژنتیکی و تحلیل حساسیت به منظور بهبود مدل‌سازی‌های مکانی در محیط GIS (مطالعه موردی: تهیه نقشه پتانسیل معدنی)

محمد اصلانی^{۱*}، محمد سعدی مسگری^۲، حمید مطیعیان^۱

^۱ دانشجوی دکتری سیستم‌های اطلاعات مکانی - دانشکده مهندسی نقشه‌برداری - دانشگاه صنعتی خواجه نصیرالدین طوسی
maslani@sina.kntu.ac.ir

^۲ دانشیار گروه سیستم‌های اطلاعات مکانی - دانشکده مهندسی نقشه‌برداری - دانشگاه صنعتی خواجه نصیرالدین طوسی
(عضو قطب علمی مهندسی فناوری اطلاعات مکانی)
mesgari@kntu.ac.ir

(تاریخ دریافت مرداد ۱۳۹۳، تاریخ تصویب مرداد ۱۳۹۴)

چکیده

یکی از مولفه‌های اصلی سامانه‌های اطلاعات مکانی مدل‌سازی‌های مکانی است. روش‌های هوشمند همانند شبکه‌های عصبی از جمله روش‌هایی هستند که در دهه‌های اخیر جهت انجام مدل‌سازی‌های مکانی به آنها توجه زیادی شده است. در اغلب روش‌های هوشمند، مدل‌سازی مکانی بر مبنای تعیین یک ساختار اولیه برای روش مورد استفاده انجام می‌شود که تعیین چنین ساختار اولیه‌ای در اغلب مواقع امری بسیار دشوار است. در این تحقیق برای حل چنین چالشی برنامه نویسی ژنتیکی مورد استفاده قرار گرفت. در این روش روابط بین داده‌های ورودی و خروجی به طور خودکار و بدون نیاز به تعریف یک ساختار اولیه و به صورت کاملاً صریح استخراج می‌شوند. یکی از چالش‌ها در بکارگیری برنامه نویسی ژنتیکی انطباق بیش از حد می‌باشد. در این تحقیق روشی نوین برای جلوگیری از مشکل انطباق بیش از حد ارائه شده است. همچنین برای افزایش دقت مدل‌سازی از تلفیق خطی وزندار چند درخت (برنامه نویسی ژنتیکی چند ژنه) استفاده شد. همچنین برای تعیین موثرترین فاکتور ورودی تاثیر گذار از دو روش تحلیل حساسیت EFAST و Sobol's که بر مبنای شبیه‌سازی مونت کارلو هستند استفاده گردید. روش پیشنهادی دارای کاربردهای متفاوتی می‌باشد و در این تحقیق از روش پیشنهادی برای تهیه نقشه پتانسیل معدنی استفاده شد. در نهایت مشخص گردید که معیارهای ژئوفیزیکی و از میان آنها فاکتور شدت میدان مغناطیسی دارای بیشترین حساسیت و معیارهای زمین‌شناسی و از میان آنها سنگ‌شناسی دارای کمترین حساسیت بر روی خروجی هستند.

واژگان کلیدی: برنامه نویسی ژنتیکی چند ژنه، تحلیل حساسیت، نقشه پتانسیل معدنی، سامانه‌های اطلاعات مکانی

۱- مقدمه

پیشرفت سامانه های اطلاعات مکانی در سال های اخیر منجر به پیشرفت های شگرفی در روش های تحلیل و مدلسازی داده های مکانی شده است [۱]. در مدلسازی های مکانی روابط بین لایه های ورودی تحلیل و استخراج می شوند تا در نهایت با بکارگیری روابط استخراج شده نقشه های خروجی بدست آیند. پیچیدگی بالای پدیده های مکانی و واضح نبودن روابط بین لایه های مکانی منجر به بکارگیری روش های پیشرفته تر از جمله روش های یادگیری ماشین در تحلیل های مکانی شده است [۳،۲]. روش های یادگیری ماشین امکان مدلسازی های مکانی را بدون نیاز به دانش کارشناسی فراوان از پدیده مورد نظر فراهم می کند. هر یک از روش های یادگیری ماشین از یک ایده خاص برای استخراج روابط میان داده ها استفاده می کنند. در اغلب این روش ها در ابتدا باید ساختار روش مورد استفاده تعیین شود (به عنوان مثال برای شبکه عصبی تعداد نرون ها و تعداد لایه های مخفی) و در ادامه پارامترها تقریب زده شوند. در صورتی که نتایج بدست آمده مطلوب نباشند ساختار اولیه آنقدر تغییر داده می شود تا نتایج مطلوب حاصل گردد [۴]. در این تحقیق برای عدم مواجه با چنین چالش هایی از برنامه نویسی ژنتیکی^۱ استفاده گردید. در این روش برخلاف سایر روش های یادگیری ماشین نیازی به تعریف ساختار اولیه نمی باشد و کلیه حالت هایی که ممکن است برای روابط میان لایه های مکانی اتفاق افتد در یک روند تکاملی مورد بررسی قرار می گیرند. در برنامه نویسی ژنتیکی رابطه بین ورودی ها و خروجی ها به صورت کاملاً صریح بدست می آید [۵]. برنامه نویسی ژنتیکی در حوزه سامانه های اطلاعات مکانی در تحقیقات محدودی از جمله پیش بینی میزان بارش [۶]، میزان تصادفات [۷]، قیمت مسکن [۸]، مدلسازی تغییر کاربری [۹] مورد استفاده قرار گرفته است.

در سال ۲۰۱۲، کاشید^۲ و مایتی^۳ از برنامه نویسی ژنتیکی برای پیش بینی میزان بارش ماهیانه در مناطق موسومی هند استفاده کردند. ایشان از الگوهای چرخش

اتمسفیری هوا برای دو اقیانوس آرام و هند به عنوان شاخص های تاثیر گذار استفاده نمودند. همچنین در زمینه پیش بینی میزان بارش با استفاده برنامه نویسی ژنتیکی می توان به تحقیقات دیگری نظیر [۱۰-۱۲] اشاره نمود.

در سال ۲۰۰۹، داس^۴ و عبدالاتی^۵ از برنامه نویسی ژنتیکی خطی برای مدلسازی رابطه بین فاکتورهای هندسی راه ها و میزان تصادفات در خیابان های ایالت فلوریدا استفاده نمودند. ایشان ۵۸ فاکتور ورودی را در مدلسازی خود لحاظ نمودند و در نهایت مشخص گردید که عدم دید کافی بیشترین تاثیر را در تصادفات منجر به جرح دارا است [۷].

در سال ۲۰۰۴، کابودان^۶ از برنامه نویسی ژنتیکی برای پیش بینی قیمت مسکن در شهر کمبریج استفاده نمودند [۸]. در سال ۲۰۰۵، منسون^۷ از تلفیق برنامه نویسی ژنتیکی و سیستم های عامل مینا برای مدلسازی تغییر کاربری در منطقه پنینسولار مکزیک استفاده کرد [۹].

یکی از موضوعات حائز اهمیت در کیفیت عملکرد برنامه نویسی ژنتیکی، مساله انطباق بیش از حد مدل بر داده های آموزشی است [۱۳] که متأسفانه به این موضوع در اغلب تحقیقات مکانی انجام شده با استفاده از برنامه نویسی توجه زیادی نشده است. در این تحقیق روشی نوین برای جلوگیری از انطباق بیش از حد مدل بر روی داده های آموزشی ارائه شده است. همچنین در روش های رایج برنامه نویسی ژنتیکی هر کروموزوم (رابطه ریاضی) فقط شامل یک درخت است. اما در تحقیق حاضر برای بالا بردن دقت مدلسازی از برنامه نویسی ژنتیکی چند ژنه استفاده گردید. بدین صورت که هر کروموزوم شامل تلفیق خطی وزن دار چندین درخت است.

یکی دیگر از نقاط قوت این تحقیق بکارگیری همزمان چندین اپراتور ادغام و برش برای افزایش تنوع پذیری کروموزوم ها جهت رسیدن به بهینه کلی می باشد. در نهایت برای انتخاب بهینه ترین رابطه از بهینه سازی چند هدفه (پرتو) استفاده گردید.

^۴ Das
^۵ Abdel-Aty
^۶ Kaboudan
^۷ Manson

^۱ Genetic Programming
^۲ Kashid
^۳ Maity

متفاوت است و در طول روند تکامل (استفاده از عملگرهای ژنتیکی) تغییر می کنند.

هر درخت از مجموعه ای از توابع و ترمینال ها تشکیل شده است. ترمینال ها مشخص کننده متغیرها و ثابت ها هستند در حالی که توابع شامل عملگرهای ریاضی هستند که بر روی متغیرها، ثابت ها و یا عملگرهای ریاضی دیگر اعمال می شوند. از جمله این توابع می توان به $\{+, /, *, -, \sin, \cos, \tan, \exp, \tanh, \sinh, \cosh, \ln\}$ اشاره نمود. توابع نیز به تعداد آرگومان های خود گره می گیرند. دلیل انتخاب توابع فوق رایج بودن آنها می باشد. فضای جستجو نیز شامل کلیه درخت های ممکن است که از تلفیق توابع و ترمینال های تعریف شده ایجاد می شوند. لازم به ذکر است، در صورتی که تعدادی از این توابع برای برقراری ارتباط بین داده های ورودی و خروجی مناسب نباشند در طول تکامل برنامه نویسی ژنتیکی به صورت خودکار حذف می شوند. مراحل برنامه نویسی ژنتیکی در ادامه شرح داده شده است.

ایجاد جمعیت اولیه: برای ایجاد جمعیت اولیه از سه روش رشدی، کامل و شبیدار نصف-نصف استفاده می شود. در روش رشدی درختان می توانند هر عمقی بین ۱ تا m (ماکزیمم عمق درخت) داشته باشند. اما در روش کامل تمام درختان کامل هستند و دارای عمق m می باشند. به عنوان یک مقایسه می توان گفت که در روش رشدی تنوع زیادی در ساختار کروموزوم ها ایجاد می شود که این موضوع در روش کامل صادق نمی باشد. در روش شبیدار نصف-نصف نیز از هر دو روش کامل و رشدی استفاده می شود و به این ترتیب تنوع ساختار و تنوع اندازه کروموزوم ها افزایش می یابد. در این تحقیق از روش نصف-نصف به دلیل تنوع پذیری بالا استفاده شد.

ارزش دهی: بعد از تعریف جمعیت اولیه ارزش هر فرد با استفاده از تابع مناسبت^۳ تخمین زده می شود. تعریف تابع مناسبت در برنامه نویسی ژنتیکی از حساسیت بسیار بالاتری نسبت به الگوریتم ژنتیک برخوردار است و انتخاب غیر دقیق آن می تواند منجر به انطباق بیش از حد شود. در این تحقیق یک تابع مناسبت نوین برای ارزیابی

به منظور انجام یک تصمیم گیری قابل اطمینان، باید بتوان میزان تاثیر هر یک از پارامترهای ورودی مدل را بر روی خروجی تعیین نمود. برای رسیدن به این مقصود از دو روش تحلیل حساسیت Sobol's و EFAST^۱ استفاده گردید که هر دوی این روش ها بر اساس شبیه سازی مونت کارلو^۲ استوار هستند [۱۴، ۱۵]. در نوشتار حاضر، کوشش شده است روش ارائه شده در یک کاربرد نمونه ارزیابی شود. بدین منظور، تهیه نقشه پتانسیل معدنی به عنوان کاربرد مطلوب در نظر گرفته شد. نقشه پتانسیل معدنی می تواند به عنوان مدلی برای تعیین نقاط بهینه حفاری چاه های اکتشافی مورد استفاده قرار گیرد.

۲- بیان روش

۲-۱- برنامه نویسی ژنتیکی

برنامه نویسی ژنتیکی یکی از الگوریتم های تکاملی است که بر پایه تئوری داروین بنا شده است و هدف از آن استخراج رابطه ریاضی تحلیلی است که به بهترین شکل ارتباط بین داده های ورودی و خروجی را برقرار سازد. مزیت برنامه نویسی ژنتیکی بر روش های رایج یادگیری ماشین نظیر شبکه های عصبی، رگرسیون و سیستم های استنتاج فازی در این است که در این روش نیازی به تعیین یک ساختار اولیه در ابتدا نمی باشد. به عنوان مثال در رگرسیون نیاز است که ساختار رابطه ریاضی (خطی، درجه دوم، چند جمله ای) در ابتدا مشخص باشد و در روند محاسبات فقط ضرائب عددی محاسبه می شود. در حالی که در برنامه نویسی ژنتیکی هم ساختار رابطه ریاضی و هم ضرائب عددی به طور مستقیم محاسبه می شوند. همچنین در برنامه نویسی ژنتیکی بر خلاف شبکه عصبی که به صورت جعبه سیاه عمل می کند رابطه ریاضی بین ورودی ها و خروجی ها به صورت کاملاً شفاف و صریح بدست می آید [۵].

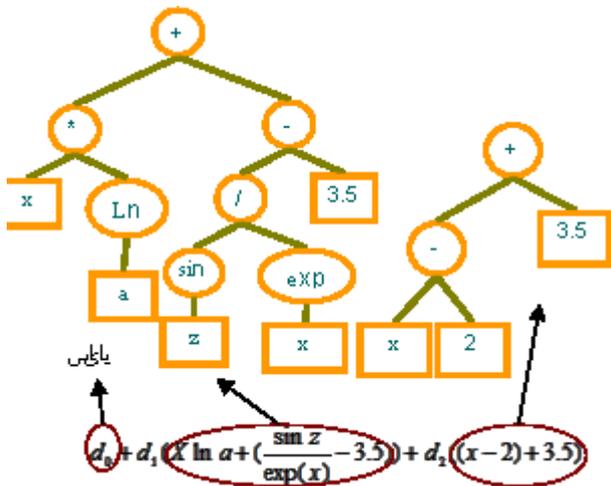
در برنامه نویسی ژنتیکی هر فرد (کروموزوم) دارای یک ساختار درختی است که نماینده یک رابطه ریاضی می باشد. اندازه و شکل درختان (کروموزوم ها) با یکدیگر

^۳ Fitness

^۱ Extended Fourier Amplitude Sensitivity Test

^۲ Monte Carlo

چند ژنه استفاده گردید. در این روش هر کروموزوم شامل تلفیق خطی وزن دار چندین درخت می باشد به گونه ای که می توان هر درخت را به عنوان یک ژن سطح بالا در نظر گرفت. شکل ۱ یک مدل چند ژنه را نشان می دهد.



شکل ۱- مدل چند ژنه

همانطور که مشخص است این مدل چند ژنه شامل تلفیق خطی عبارات های غیر خطی می باشد. مزیت این روش این است که به طور همزمان از توانایی های رگرسیون خطی و غیر خطی استفاده می کند. برای پیاده سازی این روش علاوه بر مشخص نمودن ماکزیمم عمق هر درخت، ماکزیمم تعداد ژن ها نیز باید مشخص شوند (G_{max}). ضرائب خطی (d_0, d_1, d_2, \dots) نیز به صورت مستقل و با استفاده از روش کمترین مربعات بدست می آیند. در این روش دو نوع برش قابل تعریف می باشد: ۱- برش سطح بالا و ۲- برش سطح پایین. عملکرد برش سطح پایین همانند روش های برش رایج در برنامه نویسی ژنتیکی است اما در برش سطح بالا درختی از یک کروموزوم با درختی از کروموزوم دیگر جایگزین می شود.

۲-۲- تحلیل حساسیت

هدف تحلیل حساسیت تعیین میزان تاثیر هر یک پارامترهای ورودی مدل بر روی خروجی مدل است. به عبارت بهتر تحلیل حساسیت مشخص می کند که هر یک از منابع عدم قطعیت (ورودی و مدل) چه مقدار بر

همزمان دقت و نیز جلوگیری از انطباق بیش از حد ارائه شده است.

انتخاب: عملگر انتخاب، نیروی جلوبرنده کلیه الگوریتم های تکاملی است. کلیه روش های انتخاب در الگوریتم های تکاملی، قابل استفاده در برنامه نویسی ژنتیکی هستند. اما استفاده مستقیم از آنها در برنامه نویسی ژنتیکی منجر به کاهش کارایی این روش می شود. در این تحقیق از روش انتخاب Lexicographic parsimony pressure استفاده گردید [۱۶]. این روش انتخاب تا حدود زیادی مانع از رشد غیر ضروری ابعاد درختان و در نتیجه افزایش پیچیدگی درختان می شود. به عبارت بهتر در این روش از بین درختانی که مقدار ارزش آنها با هم برابر است درختانی انتخاب می شوند که دارای ابعاد کوچکتری هستند. همچنین این روش در جلوگیری از انطباق بیش از حد تاثیر گذار است.

عملگرهای ژنتیکی: همانند سایر روش های تکاملی در برنامه نویسی ژنتیکی از عملگرهای باز تولید، جهش و برش برای تولید نسل بعدی استفاده می شود. اما با توجه به ساختار درختی که در برنامه نویسی ژنتیکی وجود دارد استفاده از عملگرهای جهش و برش رایج امکان پذیر نمی باشد و نیاز به تعریف عملگرهای جدید می باشد. برای اعمال عملگر برش در گام اول کلیه گره های درختان شماره گذاری می شوند که در این تحقیق روش عمق اول با شروع از چپ مورد استفاده قرار گرفت. در مرحله دوم دو گره به تصادف انتخاب می شوند و زیر درخت های ایجاد شده با یکدیگر تعویض می شوند.

اپراتورهای جهش در برنامه نویسی ژنتیکی از تنوع بسیار زیادی برخوردار هستند. در این تحقیق از عملگرهای جهش رشدی و ترمینال استفاده شد. در عملگر جهش رشدی به تصادف یک نود انتخاب می شود و یک زیر درخت که به تصادف ایجاد شده است جایگزین آن می شود و در عملگر جهش ترمینال، مقدار یک ترمینال به تصادف تغییر می کند. تولید نسل در برنامه نویسی ژنتیکی آنقدر ادامه می یابد تا شرط خاتمه (رسیدن به ماکزیمم تعداد نسل، کمترین مقدار تابع مناسبت) برقرار شود. در روش های رایج برنامه نویسی ژنتیکی هر کروموزوم فقط شامل یک درخت است که بیانگر یک رابطه ریاضی می باشد. اما در این تحقیق به منظور بالا بردن دقت مدلسازی از برنامه نویسی ژنتیکی

استفاده شد. در روش EFAST با استفاده از تبدیل فوریه داده ها از فضای n بعدی به فضای یک بعدی انتقال می یابند و در نتیجه انتگرال های یک بعدی استفاده می شوند. اما در روش Sobol's محدوده فاکتورهای ورودی در بازه صفر و یک نرمال می شوند و انتگرال های n بعدی بکار گرفته می شوند. تحلیل حساسیت گروهی متغیر های ورودی از جمله مزیت های روش EFAST است [۱۵].

۳- مورد مطالعاتی

برای بررسی و ارزیابی روش پیشنهادی در این تحقیق، این روش برای تهیه نقشه پتانسیل معدنی به کار گرفته شد. نقشه پتانسیل معدنی به عنوان مدلی از شکل و موقعیت توده های معدنی، دارای اهمیت ویژه ای در مطالعات اکتشاف تفصیلی است و می تواند به انتخاب محل هایی به عنوان هدف برای حفاری به منظور دست یافتن به ذخایر معدنی منجر شود. منطقه مورد مطالعه در این تحقیق، کانسار مس علی آباد با مختصات جغرافیایی $31^{\circ}39'$ شمالی و $53^{\circ}51'$ شرقی در استان یزد و ۵۵ کیلومتری جنوب غربی شهر یزد واقع شده است (شکل ۲). برای تهیه نقشه پتانسیل معدنی اصولی ترین روش استفاده از لایه های اطلاعاتی تاثیر گذار در تشکیل کانسار مس است که در این تحقیق از لایه های زمین شناسی و ژئوفیزیکی استفاده شد. تمامی لایه های مورد استفاده از شرکت ملی صنایع مس ایران اخذ شدند. لایه زمین شناسی با مقیاس ۱:۱۰۰۰ دارای تمامی ساخت ها، آلتراسیون ها و تیپ های سنگ شناسی مربوطه به همراه موقعیت دقیق و جهت حرکت گسل ها می باشد. اطلاعات مورد نیاز از نقشه زمین شناسی به سه لایه گسل ها، آلتراسیون و سنگ شناسی جداسازی شدند. لایه آلتراسیون (دگرسانیها) یک لایه برداری سطحی است که شامل سه نوع دگرسانی فیلیک، آرژیلیک و پروپیلیتیک است.

روی عدم قطعیت خروجی تاثیر می گذارد [۱۷، ۱۸]. روش های مختلفی از جمله موریس^۱، تحلیل رگرسیون و واریانس مبنا برای انجام تحلیل حساسیت ارائه شده اند که اغلب آنها بر اساس شبیه سازی مونت کارلو هستند. روش مونت کارلو امکان بررسی بازه ای تغییرات را در فاکتورهای ورودی می دهد و نیاز به داشتن فرضیات درباره ساختار مدل ندارد. در مونت کارلو فاکتورهای ورودی به عنوان متغیرهای تصادفی با تابع توزیع احتمال مشخص در نظر گرفته می شوند و بر اساس آن ها، مجموعه ای از فاکتورهای ورودی به صورت تصادفی ایجاد و خروجی متناظر با آنها تولید می شود. در این تحقیق از روش های واریانس مبنا برای انجام تحلیل حساسیت استفاده شده است. در نظر گرفتن تعاملات بین ورودی ها و مستقل از مدل بودن، جزء مزیت های روش های واریانس مبنا است. در روش های واریانس مبنا برای هر فاکتور ورودی یک شاخص حساسیت محاسبه می شود که نشان دهنده درصد مشارکت آن فاکتور ورودی در واریانس خروجی مدل است [۱۹].

مدل $y = f(x_1, x_2, \dots, x_n)$ را در نظر بگیرید. مقدار $V_i = \text{Var}(E[Y|x_i])$ نشان دهنده واریانس و نشان دهنده امید ریاضی است) بیان کننده حساسیت خروجی مدل Y به ورودی x_i است. $S_i = \frac{V_i}{\text{Var}(Y)}$ نشان دهنده شاخص حساسیت متغیر x_i است. به S_i شاخص حساسیت مرتبه اول نیز گفته می شود زیرا فقط تغییرات خروجی را نسبت به متغیر x_i می سنجد و تعامل بین چندین متغیر ورودی را در نظر نمی گیرد. شاخص حساسیت مرتبه دوم نیز به صورت زیر محاسبه می شود:

$$S_{ij} = \frac{\text{Var}(E[Y | x_i, x_j])}{\text{Var}(Y)} \quad (1)$$

شاخص حساسیت مجموع نیز عبارت است از مجموع شاخص های حساسیت مرتبه اول، دوم، سوم، ... و nام. در این تحقیق شاخص های حساسیت مرتبه اول و مجموع محاسبه شدند. روش های مختلفی برای تخمین شاخص های حساسیت مرتبه اول و مجموع وجود دارند که در این تحقیق از روش های EFAST و Sobol's

^۱ Morris



شکل ۲- منطقه مورد مطالعه

یابد. به عبارت بهتر انتخاب غیر بهینه تعداد ژن ها می تواند منجر به عدم همگرایی برنامه نویسی ژنتیکی به بهینه سراسری شود. همچنین حداکثر عمق درختان (m) نیز پارامتر حائز اهمیت دیگری است که باید مشخص شود. انتخاب کوچک این پارامتر می تواند منجر به همگرایی برنامه نویسی ژنتیکی به رابطه ساده اما غیر دقیق شود. در این تحقیق تعداد ژن های هر کروموزوم بین ۱ تا ۶ و حداکثر عمق درختان بین ۱ تا ۸ تغییر داده شدند.

در پیاده سازی برنامه نویسی ژنتیکی نرخ جهش برابر ۰,۲، نرخ برش برابر ۰,۷۵ و نرخ انتقال نخبگان نیز برابر ۰,۰۵ در نظر گرفته شدند. حداکثر تعداد نسل ها نیز برابر با ۳۰۰ نسل و هر نسل دارای ۳۰۰ کروموزوم می باشد. برای تولید جمعیت اولیه نیز از روش نصف-نصف شبیدار استفاده شد. یکی از چالش های بسیار بزرگ در برنامه نویسی ژنتیکی انتخاب تابع مناسب می باشد و انتخاب نادرست تابع مناسب می تواند منجر به انطباق بیش از حد یا عدم تکامل مناسب شود. یکی از روش های رایج برای جلوگیری از انطباق بیش از حد استفاده از داده های ارزیاب و توقف الگوریتم هنگام شروع انطباق بیش از حد است. اما استفاده از این روش با توجه به ذات تکاملی تصادفی بودن برنامه نویسی ژنتیکی ممکن است مانع از همگرایی به بهینه سراسری شود. به عنوان مثال فرض نمایید در نسل n ام انطباق بیش از حد اتفاق افتد اما با توجه به ماهیت تکاملی تصادفی برنامه نویسی ژنتیکی ممکن است در نسل k ام ($k > n$) به دلیل استفاده از عملگر جهش کروموزومی ایجاد شود که دارای کمترین میزان

لایه سنگ شناسی نیز که یک لایه برداری سطحی است شامل دایک ها، سنگ میزبان (گرانیت) و سنگ های دیواره (توف و QSC^۱) و سایر سنگ ها می باشد. مطالعات ژئوفیزیکی شامل پنج نقشه مختلف در مقیاس ۱:۲۰۰۰ با فرمت TIN^۲ است که این نقشه ها عبارتند از: شدت کل میدان مغناطیسی، مشتق دوم شدت کل میدان مغناطیسی، شارژ ابلیتته، مقاومت ظاهری و فاکتور فلزی. همچنین نقشه آنومالی ژئوشیمیایی دارای مقیاس ۱:۵۰۰۰ با فرمت TIN می باشد. با بکارگیری آنالیزهای مکانی، نقشه های اولیه به نقشه های موضوعی مناسب برای آنالیز با فرمت رستری با اندازه پیکسل یک متر تبدیل شدند. ویژگی مهم کانسار مس که آن را برای این تحقیق مناسب می سازد، وجود ۲۸ حلقه چاه اکتشافی است. متوسط عیار عنصر مس نسبت به عمق به عنوان میزان مطلوبیت چاه های اکتشافی در نظر گرفته شد. مقادیر نقشه های فاکتور در محل های چاه های اکتشافی به عنوان ورودی ها و میزان درصد خلوص مس به عنوان خروجی بردارهای آموزشی و کنترل در نظر گرفته شدند.

۴- نتایج پیاده سازی

در برنامه نویسی ژنتیکی تعداد ژن های (درخت ها، T) هر کروموزوم که تاثیر زیادی بر روی عملکرد برنامه نویسی ژنتیکی می گذارند باید از ابتدا مشخص باشند. هرچه تعداد ژن ها کمتر باشد پیچیدگی کروموزوم ها نیز کاهش می

^۱ Quartzitic Sandstone & Conglomerate

^۲ Triangulated Irregular Network

عمق درختان تشکیل دهنده کروموزوم بیشتر باشد پیچیدگی آن کروموزوم نیز افزایش می یابد. برای جلوگیری از افزایش پیچیدگی درختان از روش انتخاب Lexicographic parsimony pressure استفاده شده است. در این روش در صورتی که دو کروموزوم دارای میزان تناسب برابر یا نزدیک به هم باشند کروموزومی انتخاب می شود که دارای پیچیدگی کمتری باشد. جدول ۱ مقادیر مختلف $SAE_{Train \& \ Validation}$ و پیچیدگی را به ازای مقادیر مختلف T و m نشان می دهد. در هر یک از خانه های جدول ۱ عددی که در ابتدا آمده است بیانگر $SAE_{Train \& \ Validation}$ و عدد دوم بیانگر میزان پیچیدگی (تعداد گره ها در درختان تشکیل دهنده کروموزوم) است. همانطور که مشخص است به ازای $T=1$ و $m=2$ رابطه کروموزوم با کمترین پیچیدگی و به ازای $T=5$ و $m=6$ کروموزوم با بیشترین پیچیدگی بدست می آید. همچنین به ازای $T=5$ و $m=5$ کروموزوم با کمترین مقدار $SAE_{Train \& \ Validation}$ بدست می آید. مسلماً رابطه ای بهینه ترین رابطه است که دارای کمترین پیچیدگی و کمترین میزان خطا باشد. بنابراین انتخاب بهینه ترین رابطه از جدول ۱ یک مساله بهینه سازی چند هدفه است. شکل ۳ جبهه پرتو را به ازای دو معیار خطا و پیچیدگی نشان می دهد. نقاطی که با رنگ قرمز مشخص شده اند پاسخ های مغلوب و نقاطی که با رنگ سبز مشخص شده اند پاسخ های غیر مغلوب هستند. همانطور که از جبهه پرتو مشخص است در صورتی که فقط معیار خطا لحاظ شود رابطه ریاضی بدست آمده از برنامه نویسی ژنتیکی به ازای $T=5$ و $m=5$ دارای کمترین میزان خطا است (۶۶,۸۴۷۸). شکل ۴ میزان خطای برآورد داده های آموزشی و ارزیاب را در طول روند تکامل در نسل های مختلف به ازای $T=5$ و $m=5$ نشان می دهد. لازم به ذکر است که تابع تناسب بکار رفته برای تکامل برنامه نویسی ژنتیکی مجموع قدر مطلق خطا برای داده های آموزشی (SAE_{Train}) می باشد و از داده های ارزیاب هیچ گونه استفاده ای نشده است. روند نزولی نمودار آبی رنگ این موضوع را به وضوح نشان می دهد. همچنین نمودار سبز رنگ بیانگر مجموع قدر مطلق خطا برای داده های ارزیاب ($SAE_{Validation}$) و نمودار قرمز رنگ بیانگر $SAE_{Train \& \ Validation}$ می باشد. همانطور که از شکل ۴ مشخص است کمترین میزان $SAE_{Train \& \ Validation}$ در نسل ۱۷۸ اتفاق افتاده است. همچنین در نسل های

خطا برای داده های آموزشی و ارزیاب باشد. از اینرو استفاده از روش های رایج ممکن است مانع از رسیدن الگوریتم به بهینه سراسری شود. در این تحقیق روشی نوین برای حل چالش فوق ارائه شده است. در این روش به الگوریتم اجازه داده می شود تا به تعداد نسل های کافی (در این تحقیق ۳۰۰ نسل) با استفاده از تابع تناسب زیر تکامل یابد.

$$Fitness = SAE_{Train} = \sum_{i=1}^{N_{Train}} |y_{o_i} - y_{c_i}| \quad (2)$$

در رابطه ۲، y_{o_i} مقدار خروجی مطلوب برای داده i ام و y_{c_i} مقدار خروجی محاسبه شده برای داده i ام است و N_{Train} نیز تعداد داده های آموزشی است. بعد از خاتمه عملکرد الگوریتم، بهترین کروموزوم در هر نسل که کمترین مقدار را در رابطه ۳ داشته باشد انتخاب می شود.

$$SAE_{Train\&Validation} = W_1 \times SAE_{Validation} + W_2 \times |SAE_{Train} - SAE_{Validation}|$$

$$SAE_{Train} = \sum_{i=1}^{N_{Train}} |y_{o_i} - y_{c_i}| \quad (3)$$

$$SAE_{Validation} = \sum_{j=1}^{N_{Val}} |y_{o_j} - y_{c_j}|$$

در رابطه ۳ N_{Train} تعداد داده های آموزشی، N_{Val} تعداد داده های ارزیاب، y_{o_i} خروجی مطلوب برای داده آموزشی i ام، y_{c_i} مقدار خروجی محاسبه شده برای داده i ام است، y_{o_j} خروجی مطلوب برای داده آموزشی j ام و y_{c_j} مقدار خروجی محاسبه شده برای داده j ام است. W_1 و W_2 نیز پارامترهای تنظیمی هستند که در این تحقیق $W_1 = 0.75$ و $W_2 = 0.25$ انتخاب شدند. در رابطه فوق در صورتی که خطای برآورد داده های آموزشی (SAE_{Train}) بسیار کم و خطای برآورد داده های ارزیاب ($SAE_{Validation}$) بسیار زیاد باشد (انطباق بیش از حد) مقدار $SAE_{Train\&Validation}$ افزایش می یابد و بنابراین شانس انتخاب شدن آن کروموزوم به عنوان کروموزوم بهینه کاهش می یابد.

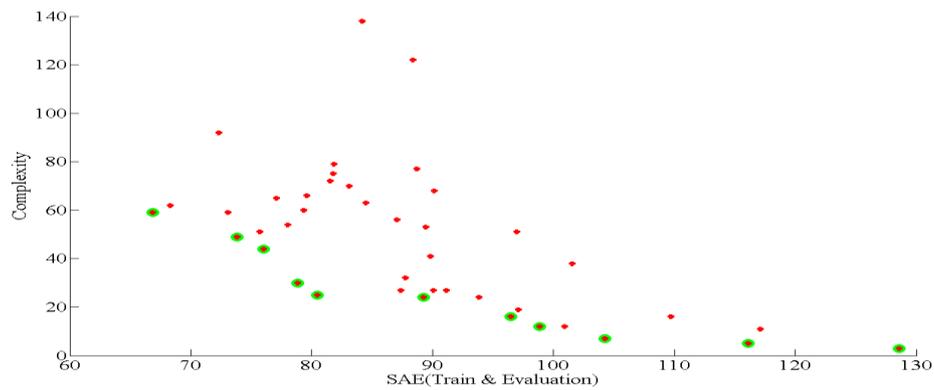
مساله دیگری که در پیاده سازی برنامه نویسی ژنتیکی باید به آن توجه شود پیچیدگی کروموزوم ها است. هرچه

شده است. شکل ۵ مقادیر خروجی مورد انتظار و محاسبه شده را برای داده های آموزشی و ارزیابی نشان می دهد.

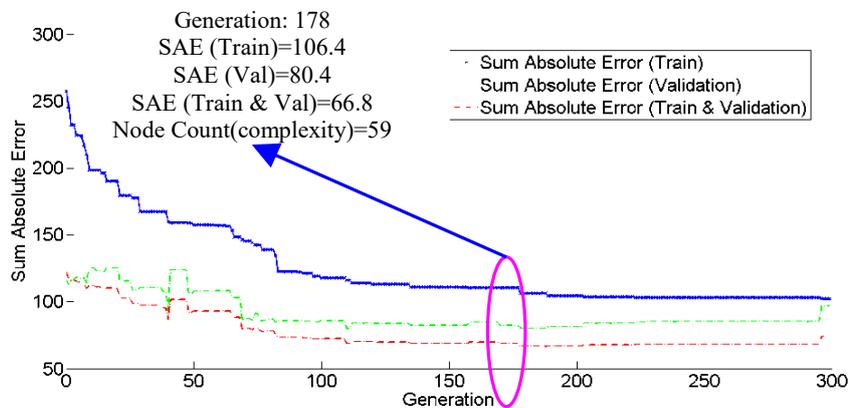
۴۰ و ۲۹۰ مقدار $SAE_{Validation}$ جهش داشته است که این موضوع بیانگر وقوع انطباق بیش از حد است که روش پیشنهادی مقاله مانع از انتخاب آنها به عنوان مقادیر بهینه

جدول ۱- مقادیر $SAE_{Train \& Validation}$ و پیچیدگی را به ازای مقادیر مختلف m و T

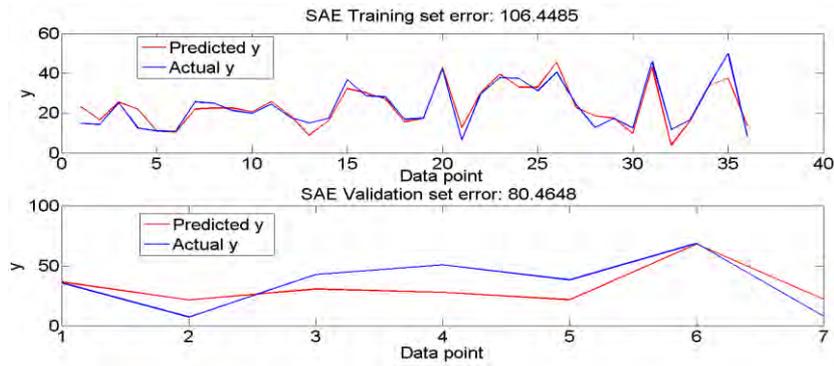
		تعداد ژن های هر کروموزوم (T)					
		۱	۲	۳	۴	۵	۶
ماکزیمم عمق درخت (m)	۲	۱۲۸,۶۰۸۹	۱۱۶,۱۱۱	۱۰۴,۲۷۰۷	۹۶,۴۹۶۵	۹۸,۸۵۹۳	۱۰۰,۹۳۴۴
		۳	۵	۷	۱۶	۱۲	۱۲
	۳	۱۱۷,۱۴۰۵	۹۷,۰۸۱۳	۸۰,۴۴۹۶	۹۳,۸۶۶۱	۹۰,۰۵۶۸	۸۷,۴۰۹۳
		۱۱	۱۹	۲۵	۲۴	۲۷	۲۷
	۴	۱۰۹,۷۳۹	۹۱,۱۴۸۷	۷۶,۰۴۴۷	۸۷,۷۶۹۴	۷۳,۰۹۴۸	۸۹,۲۳۸
		۱۶	۲۷	۴۴	۳۲	۵۹	۲۴
	۵	۸۹,۸۳۶	۷۹,۶۰۸۵	۷۵,۷۱۷۶	۹۶,۹۷۵۲	۶۶,۸۴۷۸	۸۹,۴۳۲۴
		۴۱	۶۶	۵۱	۵۱	۵۹	۵۳
۶	۸۷,۰۴۶۷	۷۹,۳۶۷	۶۸,۳۱۱۵	۷۸,۸۳۸۳	۸۴,۲۰۰۴	۷۳,۸۴۹۳	
	۵۶	۶۰	۶۲	۳۰	۱۳۸	۴۹	
۷	۸۱,۸۵۲۶	۸۸,۷۲۴۹	۷۸,۰۲۹۷	۸۸,۳۶	۷۷,۱۲۲۴	۸۳,۰۹۴	
	۷۹	۷۷	۵۴	۱۲۲	۶۵	۷۰	
۸	۷۲,۳۲۹۱	۹۰,۱۲۰۳	۸۱,۵۲۲۳	۸۴,۴۹۱۳	۸۱,۷۶۸۲	۱۰۱,۵۴۹۸	
	۹۲	۶۸	۷۲	۶۳	۷۵	۳۸	



شکل ۳- جبهه پرتو



شکل ۴- میزان خطای برآورد داده های آموزشی و ارزیابی



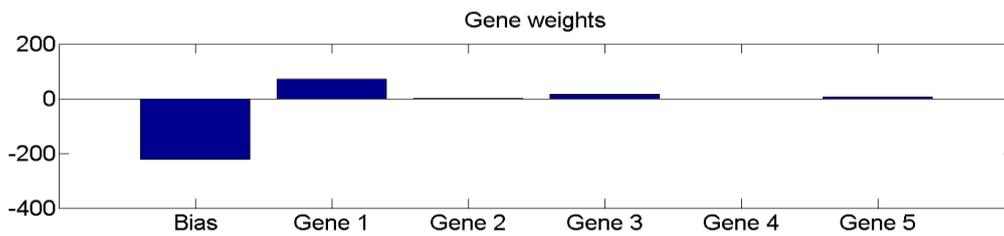
شکل ۵- خروجی مورد انتظار و محاسبه شده برای داده های آموزشی و ارزیابی

$$\begin{aligned}
 y = & -222.5 + 71.64 \operatorname{psqroot}(\operatorname{plog}(\operatorname{square}(x_4) + \sin(x_4))) + 0.6956 \cos(\operatorname{pdivide}(\cosh(x_4), x_5)) + \\
 & 0.6956 \cos(\operatorname{pdivide}(x_4, 1.759508)) + 0.6956 \sin(x_5) + 0.6956 x_5 + 16.08 \cos(\operatorname{pdivide}(x_3 + x_5, x_2)) + \\
 & 16.08 e^{\sin(x_2) \sin(x_5)} - 0.3994 \operatorname{pdivide}(\operatorname{plog}(\cosh(x_3)) - \tan(x_3 + x_6), e^{\operatorname{square}(\sin(x_5))}) + \\
 & 5.965 \operatorname{plog}(\tan(\operatorname{pdivide}(x_5, 1.722357))) + 5.965 \sin(x_3 + \sin(x_3))
 \end{aligned} \quad (4)$$

می باشد (نتایج بدست آمده از تحلیل حساسیت در جدول ۲ نیز این موضوع را تایید می کند) و در عین حال اثر آن نیز ساده نمی باشد و یک رابطه کاملاً پیچیده و غیر خطی با درصد خلوص مس دارد. مسلماً ارتباط بین لایه های زمین شناسی و درصد خلوص مس یک ارتباط پیچیده می باشد و پیچیدگی بالای رابطه ۴ نیز به روشنی گواه چنین موضوعی است.

همانطور که اشاره شد در این تحقیق از برنامه نویسی ژنتیکی چند ژنه استفاده شده است. در برنامه نویسی ژنتیکی چند ژنه هر کروموزوم از تلفیق خطی وزندار چندین درخت (ژن) تشکیل شده است. شکل ۶ وزن نهایی هر یک از ژن ها را نشان می دهد.

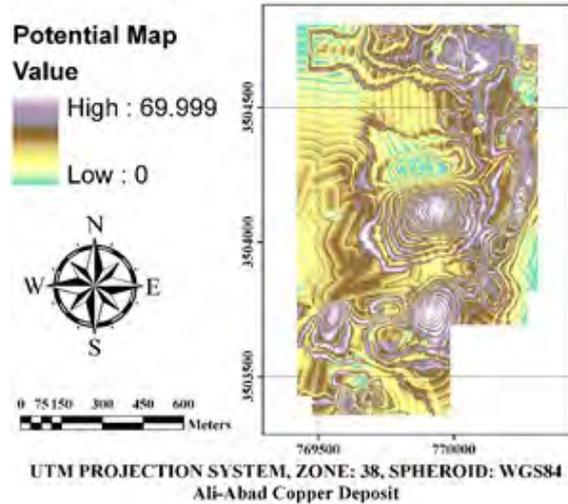
رابطه ۴ رابطه بدست آمده از برنامه نویسی ژنتیکی را نشان می دهد. در این رابطه متغیرهای X_1, X_2, X_3, X_4, X_5 و X_6 به ترتیب بیانگر معیارهای سنگ شناسی، آلتراسیون، فاصله از گسل، شارژیابلیته، شدت میدان مغناطیسی و مقاومت ظاهری هستند. همان طور که مشخص است این رابطه فاقد متغیر X_1 (معیار سنگ شناسی) می باشد بنابراین بر طبق این رابطه و در منطقه مورد مطالعه معیار سنگ شناسی نقشی در درصد خلوص مس ایفا نمی کند. همچنین متغیر X_5 (فاکتور شدت میدان مغناطیسی) در ۷ جزء از رابطه ۴ ظاهر شده است که این امر نشان می دهد که فاکتور شدت میدان مغناطیسی بیشترین حساسیت را در پتانسیل معدنی دارا



شکل ۶- وزن ژن ها

در نهایت تمامی نقشه های معیار محاسبه شده از آنالیزهای GIS وارد رابطه نهایی بدست آمده از برنامه نویسی ژنتیکی شدند و مقادیر خروجی به ازای تک تک پیکسل ها محاسبه شد. شکل ۷ نقشه پتانسیل معدنی بدست آمده را نشان می دهد.

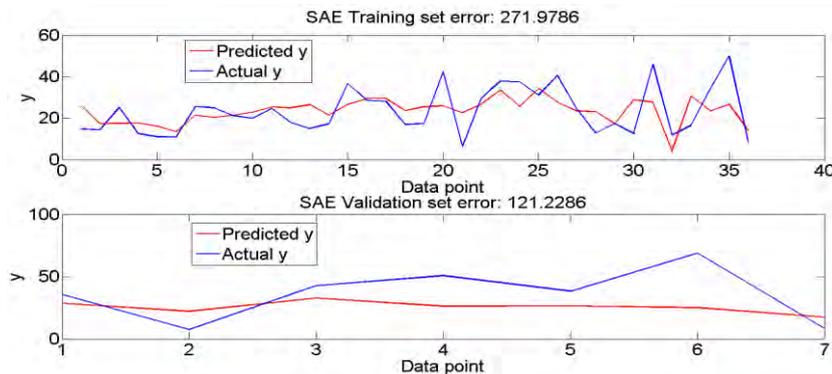
همانطور که مشخص است رابطه ۵ یک رابطه کاملاً خطی، ساده و کم دقت است. در رابطه ۵ فقط دو معیار شارژ ایلپتیه (x_4) و شدت میدان مغناطیسی (x_5) تاثیر گذار هستند و درصد اعظم تغییرات خروجی (y) توسط این دو فاکتور قابل مدلسازی است. به عبارت دیگر برای مدلسازی کم دقت پتانسیل معدنی در منطقه مورد مطالعه فقط نیاز به در نظر گرفتن دو فاکتور شارژ ایلپتیه و شدت میدان مغناطیسی است که ارزش این دو فاکتور با توجه به ضرایب یکسان آنها در رابطه ۵ با هم یکسان است. در واقع می توان انتظار داشت که در تحلیل حساسیت این دو فاکتور بیشترین ارزش را کسب نمایند. شکل ۸ مقادیر خروجی مورد انتظار و محاسبه شده را برای داده های آموزشی و ارزیابی نشان می دهد. همچنین متوسط زمان ۵ بار اجرای الگوریتم به ازای $T=1$ و $m=2$ (کمترین پیچیدگی) ۲۰۱ ثانیه و به ازای $T=5$ و $m=5$ (بیشترین دقت) ۲۹۷ ثانیه می باشد.



شکل ۷- نقشه پتانسیل معدنی

در صورتی که در جبهه پرتو تنها معیار پیچیدگی لحاظ شود بهینه ترین رابطه به ازای $T=1$ و $m=2$ بدست می آید (رابطه ۵).

$$y = -20.0 + 0.4811x_4 + 0.4811x_5 \quad (5)$$

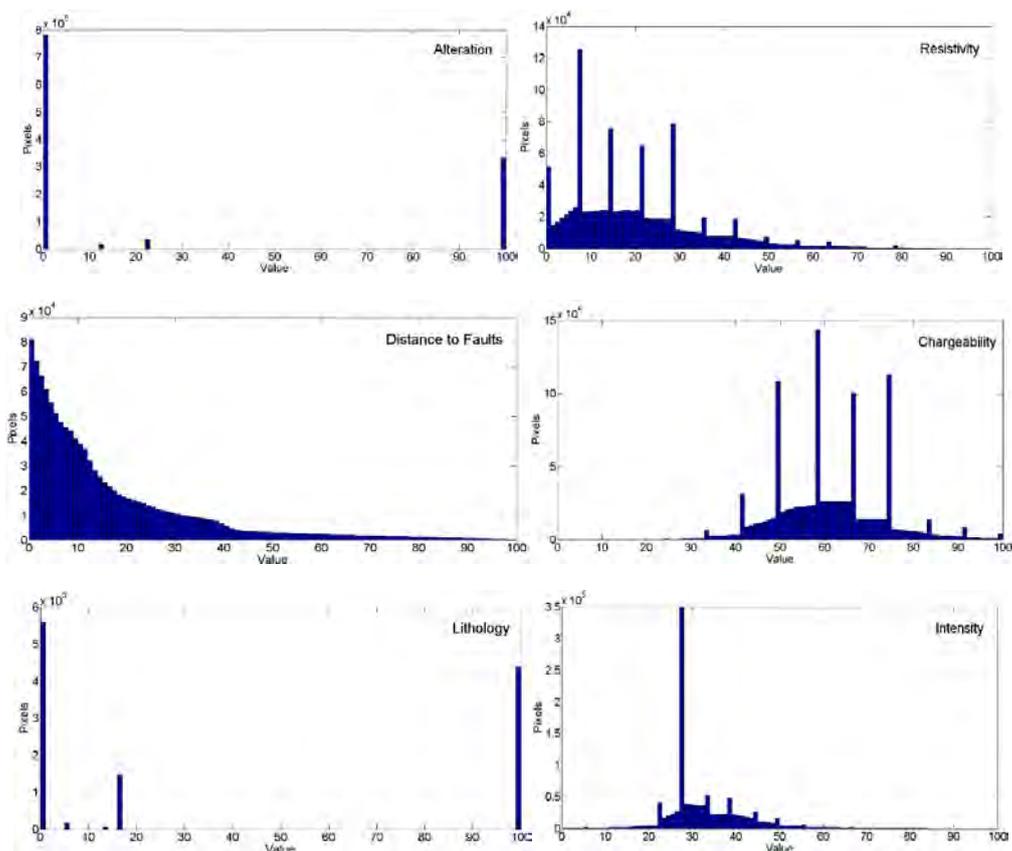


شکل ۸- مقادیر خروجی مورد انتظار و محاسبه شده برای داده های آموزشی و ارزیابی

دو روش تحلیل حساسیت Sobol's و EFAST که بر اساس شبیه سازی مونت کارلو استوار هستند مورد استفاده قرار گرفتند. در روش مونت کارلو برای هر یک از متغیرهای ورودی یک تابع توزیع احتمال مشخص در نظر گرفته شده و بر اساس آن نمونه های تصادفی تولید می شوند. در این تحقیق برای تعیین تابع توزیع احتمال فاکتورهای ورودی، از هیستوگرام نقشه های ورودی استفاده گردید [۱۸] (شکل ۹).

۵- بحث و نتیجه گیری

با پیشرفت علوم، مدل های مختلفی با پیچیدگی های مختلف برای توصیف دنیای واقعی از جنبه های مختلف توسعه داده شده اند. در استفاده از این مدل ها بررسی میزان تاثیر ورودی ها بر روی خروجی ها ضروری به نظر می رسد. روش های تحلیل حساسیت چارچوب مناسبی برای تعیین میزان اثر پذیری خروجی ها از تغییرات ورودی ها را فراهم می کنند. تحلیل حساسیت امکانات فراوانی را برای ساخت و استفاده از مدل در اختیار محققین قرار می دهند. بهبود ساختار مدل و ساده سازی مدل از جمله مزایایی استفاده از تحلیل حساسیت است.



شکل ۹- هیستوگرام نقشه های ورودی

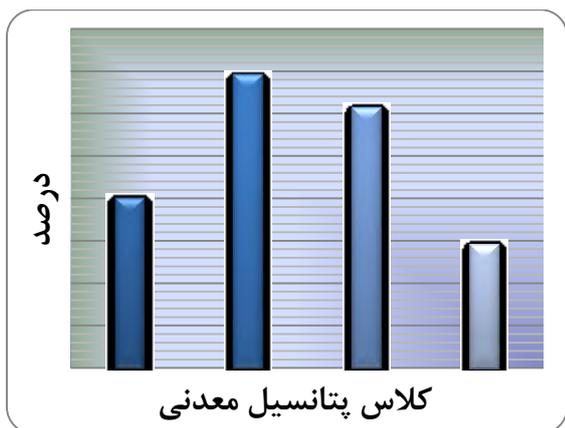
در جدول ۲ هر دو شاخص مرتبه اول و مجموع آورده شده اند. شاخص مجموع Sobol's بر خلاف شاخص مرتبه اول آن تعامل میان فاکتورها را نیز در نظر می گیرد. همانطور که مشخص است نقشه خروجی دارای بیشترین حساسیت به فاکتور شدت میدان مغناطیسی و کمترین حساسیت به فاکتور سنگ شناسی می باشد. به عبارت دیگر واریانس نقشه خروجی بیشترین تاثیرپذیری را از واریانس نقشه میدان مغناطیسی دارا است و برای رسیدن به دقت های بالاتر باید نقشه معیار شدت میدان مغناطیسی را با دقت بالایی تعیین نمود. همچنین به دلیل کمترین میزان حساسیت مدل به لایه سنگ شناسی می توان آنرا با یک مقدار میانگین بدون اینکه تغییر زیادی در نتایج حاصل شود جایگزین نمود.

روش تحلیل حساسیت EFAST نیز یکی دیگر از روش های تحلیل حساسیت است که امکان تحلیل حساسیت گروهی فاکتورها را فراهم می کند. جدول ۳ نتایج تحلیل حساسیت روش EFAST را برای شاخص مرتبه اول و کلی نشان می دهد.

برای تولید نمونه های تصادفی از روش LP_{τ} که یک روش تولید نمونه شبه تصادفی است استفاده گردید [۱۵]. در این روش نمونه ها به گونه ای تولید می شوند که تمام فضای ورودی را به صورت یکنواخت پوشش دهند و همین موضوع منجر به بهبود نتایج تحلیل حساسیت می شود. جدول ۲ نتایج تحلیل حساسیت روش Sobol's را نشان می دهد.

جدول ۲- نتایج تحلیل حساسیت روش Sobol's

Factors	Sobol First order indexes	Sobol total order Indexes
سنگ شناسی	۰	۰
آلتراسیون	۰,۰۲۷۱۵۲	۰,۰۷۰۵۹۲
فاصله تا گسل	۰,۱۸۸۱۴	۰,۲۴۲۰۳۳
شارژ ایلینته	۰,۱۰۰۸۱۹	۰,۰۹۵۷۲۸
شدت میدان مغناطیسی	۰,۴۶۶۰۴۸	۰,۵۴۷۹۳۹
مقاومت ظاهری	۰,۰۲۲۵۷۹	۰,۰۴۹۶۹۷



شکل ۱۰- مقایسه درصد مساحت کلاس های مختلف

تشکر و قدردانی

از همکاری صمیمانه شرکت ملی صنایع مس ایران به دلیل در اختیار قرار دادن داده های مورد نیاز سپاس گذاری می کنیم.

جدول ۳- نتایج تحلیل حساسیت روش EFAST

Group Factor	Fast first order indexes	Fast total order Indexes
زمین شناسی	۰,۱۴۵۵	۰,۶۲۳۸۴
ژئوفیزیک	۰,۴۳۲۶	۰,۸۴۹۶۵۴

همانطور که از جدول ۳ مشخص است فاکتور ژئوفیزیکی دارای بیشترین تاثیر بر روی خروجی است. به منظور ارزیابی بیشتر، نقشه پتانسیل معدنی بدست آمده با استفاده از روش برش طبیعی که نقاط با تغییرات شدید را به عنوان حد تمایز انتخاب می کند، به چهار کلاس تقسیم بندی شدند. با بررسی مساحت هر کلاس نسبت به کل مساحت منطقه می توان راجع به نقشه های پتانسیل معدنی بدست آمده بهتر قضاوت کرد (شکل ۱۰). همانطور که مشخص است تنها ۱۵٪ از منطقه دارای پتانسیل بالایی است.

مراجع

- [1] M. F. Manfred and A. Getis,(2010) "Advances in Spatia Analysis," in Recent Developments in Spatial Analysis: Spatial Statistics, Behavioural Modelling, and Computational Intelligence (Advances in Spatial Science), M. F. Manfred and A. Getis, Eds. Heidelberg: Springer.
- [2] J. Li, A. D. Heap, A. Potter, and J. J. Daniell,(2011) Application of machine learning methods to spatial interpolation of environmental variables, Environmental Modelling & Software, Vol. 26, No. 12, pp. 1647-1659.
- [3] M. Kanevski ,R. Parkin, A. Pozdnukhov, V. Timonin, M. Maignan, V. Demyanov, and S. Canu,(2004) Environmental data mining and modeling based on machine learning algorithms and geostatistics, Environmental Modelling & Software, Vol. 19, No. 9, pp. 845-855.
- [4] J. M. Zurada,(1992) Introduction To Artificial Neural Systems. New York: West Publishing Company.
- [5] J. R. Koza, M. A. Keane, M. J. Streeter, W. Mydlowec, J. Yu, and G. Lanza,(2003) Genetic Programming IV:Routine Human-Competitive Machine Intelligence Norwell Massachusetts:Kluwer Academic Publishers.
- [6] S. S. Kashid and R. Maity,(2012) Prediction of monthly rainfall on homogeneous monsoon regions of India based on large scale circulation patterns using Genetic Programming, Journal of Hydrology, Vol. 454-455 , No. 2012, pp. 26-41.
- [7] A. Das and M. Abdel-Aty,(2010) A genetic programming approach to explore the crash severity on multi-lane roads, Accident Analysis and Prevention, Vol. 42, No. 2, pp. 548-557.
- [8] M. Kaboudan,(2004) "Spatiotemporal forecasting of housing prices by use of genetic programming," in The 16th Annual Meeting of the Association of Global Business Mexico.
- [9] S. M. Manson,(2005) Agent-based modeling and genetic programming for modeling land change in the Southern Yucatan Peninsular Region of Mexico, Agriculture, Ecosystems and Environment, Vol. 111, No. 1-4, pp. 47-62.
- [10] V. Babovic,(2005) Data mining in hydrology, Hydrological Processes, Vol. 19, No. 7, pp. 1511-1515 ,

- [11] S.-Y. Liong, T. R. Gautam, S. T. Khu, V. Babovic, M. Keijzer, and N. Muttill,(2002) Genetic Programming: A New Paradigm in Rainfall Runoff Modeling, Journal of American Water Resources Association, Vol. 38, No. 3, pp. 705-718.
- [12] P. A. Whigham and P. F. Crapper,(2001) Modelling rainfall-runoff using genetic programming, Mathematical and Computer Modelling, Vol. 33, No. 6-7, pp. 707-721.
- [13] J. R. Koza,(1992) Genetic Programming: On the Programming of Computers by Natural Selection. Cambridge: MIT Press.
- [14] A. Saltelli, M. Ratto, T. Andres, F. Campolongo ,J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola,(2008) Global Sensitivity Analysis: The Primer. Chichester: John Wiley & Sons.
- [15] A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto,(2004) Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models. Chichester: John Wiley & Sons.
- [16] S. Luke and L. Panait,(Year) "Lexicographic Parsimony Pressure," in Proceedings of the Genetic and Evolutionary Computation Conference, New York, pp. 820-828.
- [17] M. Crosetto and S. Tarantola,(2001) Uncertainty and sensitivity analysis: tools for GIS-based model implementation, International Journal of Geographical Information Science, Vol. 15, No. 5, pp. 415-437.
- [18] M. Gómez-Delgado and S. Tarantola,(2006) Global sensitivity analysis, GIS and multi-criteria evaluation for a sustainable planning of a hazardous waste disposal site in Spain, International Journal of Geographical Information Science, Vol. 20, No. 4 pp. 449-466.
- [19] L. Lilburne and S. Tarantola,(2009) Sensitivity analysis of spatial models, International Journal of Geographical Information Science, Vol. 23, No. 2, pp. 151-168.