**ارائه يك روش جديد در تشخيص خطاهاي بارز تکی و خوشه ای**

**مدل ارتفاعی رقومی زمین بر مبناي بكارگيري مفاهيم الگوريتم‌ ژنتیک**

**فرهاد صمدزادگان1، محسن گودرزی2 ، مهران قندهاری3**

1. دانشیار، گروه مهندسی نقشه برداری و ژئوماتیک، دانشگاه تهران

samadz@ut.ac.ir

2. دانشجوی کارشناسی ارشد سیستم های اطلاعات مکانی ، گروه مهندسی نقشه برداری و ژئوماتیک، دانشگاه تهران

sahebi@kntu.ac.ir

3. دانشجوی کارشناسی ارشد سیستم های اطلاعات مکانی، گروه مهندسی نقشه برداری و ژئوماتیک، دانشگاه تهران

goodarzi.64@gmail.com

چکیده:

مدل های ارتفاعی رقومی زمین یکی از مهمترین داده های مکانی در سیتم های اطلاعات مکانی هستند. مدل های رقومی زمین در مراحل نمونه برداری، اندازه گیری و درون یابی با انواع مختلفی از خطا آلوده می شوند. بنابراین دقت یکی از مشخصه های مهم آن ها به شمار می رود. خطا ها به سه گروه؛ خطاهای تصادفی، سیستماتیک و اشتباهات تقسیم می شوند. الگوریتم ارائه شده در این مقاله به دنبال کشف خطاهایی از نوع اشتباه، در مدل های ارتفاعی رقومی غیر منظم با استفاده از مفاهیم بهینه سازی در هوش مصنوعی است. . در این روش پس از تقسیم منطقه به بخش های کوچک تر، هر بخش به صورت مجزا مورد کاوش قرار می گیرد. در هر بخش با برازش یک سطح دوخطی و محاسبه ی بردار باقیمانده ها با استفاده از مفاهيم الگوريتم‌ ژنتیک سعی در کمینه کردن مجموع مربعات بردار باقیمانده ها و کشف خطاها می شود. مزیت این روش نسبت به روش های موجود، پیدا کردن خطاهای بارز تکی و خوشه ای با دقت بالا و بدون استفاده از روش های آماری است.

**واژه‌هاي كليدي:** مدل ارتفاعی رقومی زمین، خطاهاي بارز تکی و خوشه ای، سطح دوخطی ، الگوريتم‌ ژنتیک.

1- مقدمه:

مدل های رقومی زمین یکی از منابع توسعه مدل های اطلاعات مکانی و سیستم های اطلاعات مکانی (GIS) می باشد. روش های مختلفی جهت تولید مدل های رقومی زمین وجود دارد. به عنوان مثال، نقشه برداری زمینی، رادار، امواج صوتی، عکسبرداری هوایی، رقومی سازی نقشه های توپوگرافی، InSAR و LIDAR از آن جمله می باشند. کاربرد وسیع مدل های رقومی زمین در زمینه های مختلف، کاربران را بر آن می دارد که از میان انواع مدل ها، مدلی را انتخاب نمایند که بیشترین تطبیق را با زمینه کاری خود داشته باشد. یکی از مشخصات تعیین کننده در انتخاب مدل رقومی، دقت آن است که عوامل تاثیرگذار در آن به دو دسته طبقه بندی می شود. یکی دقت در داده های اصلی مانند تراکم نمونه برداری و خطاهای اندازه گیری و دیگری در روش های به کار رفته در درونیابی می باشد. خطا ها به سه دسته؛ تصادفی، سیستماتیک و اشتباه تقسیم می شوند. در بسیاری از کاربردهای سیستم های اطلاعات مکانی فرض بر نبود اشتباهات است و تنها خطاهای اتفاقی و سیستماتیک در تعیین دقت نقش دارند. این فرض به دلیل سختی تشخیص اشتباه طی مراحل کاری است. از آن جا که وجود اشتباهات می تواند منجر به نتایج دور از انتظار شود لازم است که با روش هایی این اشتباهات را تشخیص داد و از داده های دیگر جدا نمود.[6]

اولین تلاش ها توسط Hannah(1981) انجام گرفت. او به منظور تشخیص اشتباهات در مدل های رقومی منظم از مشخصه شیب و تغییر شیب در هر نقطه و مقایسه آن با نقاط همسایه استفاده کرد. Ostman (1987) اشاره می کند که هیچ معیار واحدی برای ارتقای دقت در مدل رقومی زمین وجود ندارد و حداقل دقت در ارتفاع، شیب و انحنا باید لحاظ شود. Felicisimo (1994) روشی بر مبنای اختلاف ارتفاع بین یک نقطه و ارتفاع محاسبه شده آن، از همسایگانش ارائه داد. او فرض کرد که اختلافات دارای توزیع گوس هستند و مقداری آماری از اختلافات را بر اساس توزیع t-student تنظیم نمود. نتایج آزمایشی او نشان داد که بسیاری از اشتباهات خود را با مقدار آماری بزرگ نشان می دهند. پس به طور کلی روش ها به دو دسته 1) شیب مبنا و 2) نقطه مبنا تقسیم می شوند، روش اول برای مدل های منظم و روش دوم برای مدل های غیر منظم مناسب هستند. در روش اول از شیب، تغییرات شیب و اختلاف تغییرات شیب استفاده می شود Hannah(1981). در روش دوم ارتفاع هر نقطه نسبت به همسایگانش مقایسه و نهایتا نقاط کاندید اشتباه شناسایی می شوند.[9]

در این مقاله یک روش جدید در تشخیص و حذف مشاهدات دارای خطاهای بارز بر مبنای الگوریتم های فرااکتشافی تکاملی ارائه گردیده است.

در دهه های اخیر روش های تکاملی و فرااکتشافی به عنوان یک ابزار جستجو و بهینه سازی در حوزه های مختلف مانند علوم، تجارت و مهندسی مورد استفاده قرار گرفته اند. وسعت دامنه ی کاربرد و سهولت استفاده و قابلیت دستیابی به جواب نزدیک به بهینه مطلق از جمله دلایل موفقیت این روش ها می باشد. در این روش پس از تقسیم منطقه به بخش های کوچک تر، هر بخش به صورت مجزا مورد کاوش قرار می گیرد. در هر بخش با برازش یک سطح دوخطی و محاسبه ی بردار باقیمانده ها با استفاده از روش های فرااکتشافی سعی در کمینه کردن مجموع مربعات بردار باقیمانده ها می شود.

2- مروری بر الگوریتم ژنتیک:

الگوریتم های ژنتیکی که بر مبنای ایده تکامل در طبیعت عمل می نمایند، بر روی جمعیتی1از راه حل های بالقوه به جستجوی راه حل نهایی می پردازد. در هر نسل، بهترین های آن نسل انتخاب می شوند، و پس از زاد و ولد، مجموعه جدیدی از فرزندان را تولید می کنند. در این فرایند افراد مناسبتر با احتمال بیشتری در نسل های بعدی باقی خواهند ماند. در آغاز الگوریتم، تعدادی از افراد (جمعیت اولیه) به صورت تصادفی ساخته شده و تابع هدف2برای تک تک آنها ارزیابی می شود. اگر شرط رسیدن به جواب برقرار نباشد (به جواب بهینه نرسیده باشیم)، نسل های بعدی با انتخاب والدین بر اساس میزان برازندگی شان تولید می شود و فرزندان با احتمال ثابتی دچار جهش می شوند. سپس میزان برازندگی فرزندان جدید محاسبه شده و جمعیت جدید، از جایگزینی فرزندان با والدین ایجاد می شود و این فرایند تا برقرار شدن شرط خاتمه تکرار می شود. عمده ترین مزایای این روش در مقایسه با روش های متداول عبارتند از: جستجوی موازی در عوض جستجوی ترتیبی، عدم نیاز به هر گونه اطلاعات کمکی نظیر روش حل مساله، قطعی نبودن الگوریتم، پیاده سازی آسان و رسیدن به چند گزینه مطلوب.[13]

3- متدولوژی روش پیاده سازی شده:

**3-1- ژن و کروموزوم**

در الگوریتم وراثتی هر کروموزوم بیان گر یک جواب مساله به صورت رمز شده است و یک نقطه در فضای جستجو را نشان می دهد. هر کروموزوم از تعداد مشخصی ژن تشکیل شده است که بیانگر متغیر های مساله هستند. در این جا هر کروموزوم دارای تعداد ژن ثابث (به طور پیش برابر 10) می باشد، که هر ژن نشان دهنده ی شماره ی یک نقطه است. هدف پیدا کردن کروموزومی است که فقط شامل نقاطی باشد که آن نقاط دارای خطای بارز می باشند. طول کروموزوم به گونه ای انتخاب می گردد که تعداد نقاط دارای خطای بارز کوچکتر از طول کروموزوم باشد. به همین دلیل تعدادی از ژن ها دارای مقدار صفر هستند.

**3-1- جمعیت اولیه**:

مجموعه ای از کروموزوم ها یک جمعیت را می سازند. در اینجا جمعیت اولیه به صورت تصادفی ایجاد می گردد.

**3-2- تابع شایستگی (برازندگی):**

برای حل هر مساله بهینه سازی، باید یک تابع شایستگی طراحی شود. تابع شایستگی برای هر کروموزوم، یک عدد نامنفی بر می گرداندکه نشان دهنده کارآیی یا توانایی آن کروموزوم در حل مساله است. در این مساله برای محاسبه ی ارزش هر کروموزوم به صورت زیر عمل می شود :

ابتدا نقاط موجود در کروموزوم از مجموعه ی نقاط حذف می گردد. پس از آن یک معادله ی دوخطی به مجموعه ی نقاط برازش داده می شود و مراحل زیر برای محاصبه ی ارزش کروموزوم انجام می گیرد :

معادله ی دوخطی و مجهولات آن به صورت زیر می باشد :

(1)

(2)

که برای محاصبه ی بردار مجهولات، پس از تشکیل معادلات مشاهدات داریم :

(3)

و بردار باقیمانده ها برابر است با :

(4)

و در نهایت داریم :

(5)

هدف مینیمم کردن تابع هزینه است، زیرا زمانی تابع بالا مینیمم می شود که تمام نقاط دارای خطای بارز از مجموعه ی نقاط بیرون کشیده شود. علت تقسیم کردن مجموع مربعات باقیمانده ها بر طول بردار باقیماندها به توان 10 در زیر با ذکر یک مثال توضیح داده شده است :

فرض کنید تعداد نقاط n، طول کروموزوم m و تعداد ژن های مخالف صفر در یک کروموزوم k باشد. در صورتی که تابع هزینه برابر با مجموع مربعات باقیمانده ها باشد، هرچه مقدار k به m نزدیک تر باشد، مقدار تابع هزینه کمتر می شود زیرا نقاط بیشتری از مجموعه ی نقاط بیرون کشیده شده و طول بردار باقیمانده ها کوچک تر می گردد. برای حل این مشکل مجموع مربعات باقیمانده ها بر طول بردار باقیماندها به توان 10 تقسیم شد زیرا معمولاً تعداد نقاط دارای خطای بارز در هر قطعه از نصف طول کروموزوم کمتر است. در نتیجه با استفاده از این ضریب تجربی کروموزومی که فقط شامل نقاط دارای خطای بارز است دارای مینیمم هزینه است. در مورد کشف خطاهای خوشه ای نباید از این ضریب استفاده کرد زیرا باعث عدم کشف تمامی خطاها می گردد.

**3-3- عملگرها:**

در الگوریتم وراثتی برای ساخت نسل3بعد، از تولید مثل استفاده می شود. در مرحله تولید مثل، تعدادی از افراد جمعیت نسل فعلی انتخاب می شوند. این اعضا ضمن ترکیب4با یکدیگر نسل جدید را می سازند. هر یک از اعضای انتخاب شده برای تولید مثل، والد5و هر یک از اعضای تولید شده، فرزند6نامیده می شوند. برای تولید مثل و ساختن نسل بعد از عملگرهای وراثتی استفاده می شود که مهم ترین آن ها، عملگرهای انتخاب7، همبری8و جهش9هستند.

**3-3-1- عملگر ابتکاری:**

وظیفه ی این عملگر کشف نقاط دارای خطای بارز است به همین علت مهمترین عمگر به کار رفته برای حل این مساله می باشد. هدف این عملگر ایجاد کروموزوم جدید با استفاده از بردار باقیمانده ها است. اولین ژن کروموزوم جدید برابر با شماره ی نقطه با بیشترین مقدار باقی مانده است. سایر ژن های کروموزوم نیز بر همین اساس ایجاد می گردد. به علت استفاده از تابع دوخطی باقیمانده های متناظر با نقاط دارای خطای بارز دارای اختلاف فاحش نسبت به سایر نقاط است.

**3-3-2- عملگر انتخاب**:

با استفاده از این عملگر از بین کروموزوم های موجود در یک جمعیت تعدادی برای تولید نسل انتخاب می شوند. فرایند انتخاب به گونه ای است که کروموزوم های با شایستگی بیشتر از شانس بیشتری برای انتخاب و تولید مثل برخوردار می شوند. در طی فرایند انتخاب این امکان وجود دارد که بعضی از کروموزوم های با شایستگی زیاد چندین مرتبه انتخاب شده و بعضی از کروموزوم های ضعیف حتی برای یک مرتبه هم انتخاب نشوند.

**3-3-3- عملگر همبری:**

برای ساختن نسل بعد، دو کروموزوم به عنوان والد انتخاب شده و با عمل همبری دو فرزند به وجود می آید. ساده ترین و مشهورترین نوع همبری، همبری تک نقطه ای است، که جزئیات آن در شکل (1) نشان داده شده است. محل برش به صورت تصادفی انتخاب شده است.

|  |
| --- |
|  |
| شکل 1. عملگر همبری تک نقطه ای |

**3-3-4- عملگرجهش:**

این عملگر به طور عام بعد از عملگر همبری اعمال می شود. عملگر جهش به طور عام برای تنوع بخشیدن به جمعیت و ایجاد نقاط جدید جستجو اعمال می شود. همان گونه که در شکل (2) نشان داده شده است، مقدار تعدادی از ژن ها به صورت رندوم صفر می شود.

|  |
| --- |
|  |
| شکل 2. یک کروموزوم قبل و بعد از عملگر جهش |

4- ارزیابی و تست:

به منظور ارزیابی توانایی روش ارائه شده در این نوشته، نسبت به تشخیص خطاهای بارز موجود در یک مدل ارتفاعی رقومی زمین با 20000 نقطه به صورت غیر منظم اقدام گردید.

همان گونه که در شکل (3) مشخص است منطقه به یک شبکه 10\*10 تقسیم شد و 30 نقطه به صورت رندوم انتخاب و به ارتفاع آن نقاط مقدار 50 متر اضافه گردید. پس از اجرای برنامه ی ژنتیک تعداد 28 نقطه از 30 نقطه دارای خطای بارز، تشخیص داده شد. این الگوریتم برای نقاط دارای خطای خوشه ای نیز بررسی گردید. مطابق شکل (4)، 6 گروه و در هر گروه 8 نقطه ایجاد گردید و پس از اجرای برنامه نتایج قابل قبولی گرفته شد که در نمودار (1) آورده شده است.

|  |
| --- |
|  |
| شکل 3.نتایج کشف خطای تکی به همراه نمودار همگرایی الگوریتم ژنتیک |
|  |
| شکل 4.نتایج کشف خطای خوشه ای به همراه نمودار همگرایی الگوریتم ژنتیک |
|  |
| نمودار 1.آمار حاصل از کشف خطاهای تکی و خوشه ای |

6- پیشنهادات:

در جهت بهبود نتایج الگوریتم پیشنهاد می گردد که از سایر الگوریتم های فرااکتشافی جهت کشف خطا استفاده گردد. مثلاً الگوریتم های جست و جوی محلی می توانند با سرعت و دقت بیشتری عمل کنند. در این تحقیق برای کشف تعداد دقیق نقاط دارای خطای بارز از ضریبی تجربی در فرمول (5) استفاده شد. پیشنهاد می گردد بدون استفاده از آن ضریب تعداد نقاط بیشتری را به عنوان نقاط مشکوک به اشتباه کشف واحتمال اشتباه بودن هر یک از نقاط را بر حسب درصد بیان کرد. میزان این احتمال را می توان با استفاده از باقیمانده ی هر نقطه نسبت به سایر نقاط بدست آورد.در انتها بهتر است، نقاط مشکوک با ترسیم مدل 3 بعدی بررسی گردد.

7- نتيجه‌گيري:

نتایج حاصل از تست این روش برای کشف نقاط دارای خطای بارز در قسمت قبل، نشان دهنده ی دقت بالای این روش در کشف خطاهای منفرد و خوشه ای می باشد. می توان ادعا کرد هیچ یک از روش های موجود قادر به کشف خطاها با این سطح از دقت نمی باشند. همه ی روش های موجود سعی در حل مساله با پارامترهای آماری هستند و این موضوع باعث پایین آمدن دقت کار می شود. یکی از مزایای این روش استفاده از تابع دوخطی می باشد که باعث می گردد باقیمانده ی نقاط دارای خطای بارز دارای اختلاف فاحش با سایر نقاط باشد. یکی دیگر از مزایای این روش که در عین سادگی باعث بالا بردن دقت برای کشف خطاهای خوشه ای شده است عدم توجه به همسایگی نقاط است. در واقع همه ی نقاط موجود در یک خوشه دارای باقیمانده ی بالا می باشند. سرعت روش فوق نیز باعث بالا رفتن کارایی روش فوق شده است.

8 – پانوشت:

1.Population

2.Objective function

3.Generation

4.Recombination

5.Parent

6.Offspring

7.Selection

8.Crossover

9.Mutation

مراجع:

1. A.  Felicisimo,  Parametric  statistical  method  for error detection in digital elevation models, ISPRS Journal  of  Photogrammetry  and  Remote  Sensing  49 (1994) 29.
2. C. Lopez, On the improving of elevation accuracy of  Digital  Elevation  Models:  a  comparison  of some  error  detection  procedures,  Scandinavian Research  Conference  on  Geographical  Information Science (ScanGIS), Stockholm, Sweden,

(1997) 85.

1. Felicísimo, A. 1994. Parametric statistical method for error detection in digital elevation models. ISPRS Journal of Photogrammetry and Remote Sensing, vol. 49, pp. 29-33.
2. Fisher, P. 1988. Improved Modeling of Elevation Error with Geostatistics. GeoInformatica, vol. 2, no. 3pp. 215-233
3. Forstner, W., 1982. Results of test I on grosseError detection of ISP WG III/L and OEEPE. Proceedings of ISPRS-Symposium (Commission III), Helsinki 1982.
4. Hannah, M J. 1981. Error detection and correction in Digital Terrain Models. Photogrammetric Engineering and Remote Sensing, vol. 47, no. 1, pp. 63-69.
5. Hunter, G J and Goodchild M F. 1997. Modeling the uncertainty of slope and aspect estimates derived from spatial databases. Geographical Analysis, vol. 29-1, pp. 35-49.
6. Lee, J., Snyder P K, and Fisher P F. 1992. Modeling the effect of data errors on feature extraction from digital elevation models. Photogrammetric Engineering and Remote Sensing, vol. 58-10, pp. 1461-1467.
7. Li Z, Zhu Q. 2000. Digital Elevation Model. Wuhan. (in Chinese).
8. M.  Hannah,  Error  detection  and  correction  in digital terrain models, Photogrammetric Engineering and Remote Sensing 47 (1981) 63.
9. Östman, A. 2000. Quality control of photogrammetrically sampled Digital Elevation Models. Photogrammetric Record, vol. 12, pp. 333-341.
10. Pope, A.J., 1976. The statistics of residuals and the detection of outliers. NOAA Technical Report NOS 65. NGS l, U.S. Dept. of Commerce, Rockville. Md.
11. Randy L. Haupt, Sue Ellen Haupt,Practical Genetic Algorithm,WILEY,2004
12. Tang, G. 2000. A Research on the Accuracy of Digital Elevation Model. Science Press, Beijing, New York, 2000.
13. Werner, H., 1984. Automatic Gross Error Detection by Robust Estimators, IAP-Vol. 25-III, Rio de Janeiro 1984.
14. Z.L. Li, Sampling Strategy and Accuracy Assessment for  Digital  Terrain  Modelling,  Ph.D.  thesis,  The University of Glasgow, 1990.
15. Z.L. Li, Q. Zhu, C. Gold, Digital terrain modeling: principles  and  methodology,  CRC  Press,  Boca Raton, 2005.