

APLICACIÓN DE ALGORITMOS GENÉTICOS DE CONVERGENCIA RÁPIDA EN IMÁGENES MULTIESPECTRALES DE ALTA RESOLUCIÓN ESPACIAL PARA LA EXTRACCIÓN DE CRUCES DE CAÑERÍAS SOBRE EL RÍO NEUQUÉN – YACIMIENTO LOMA LA LATA – PROVINCIA DEL NEUQUÉN – ARGENTINA

José Martín Catalini
SUMICOMP S.R.L.

Santa Fe 53 Piso 3 Dpto. 'E'
Neuquén – 8300 – Argentina
Tel: +54299-4480360

E-mail: jcatalinif@repsolypf.com – sumicomp@neuquen.com.ar

SINOPSIS

El desarrollo de actividades de producción de hidrocarburos en zonas ambientalmente sensibles es un impulso auténtico para generar actividades orientadas a diagnosticar la situación de cada área de operaciones y construir mecanismos para la identificación y prevención de incidentes potenciales. Un análisis pormenorizado de las localizaciones precisas de los cruces de cañerías con la hidrografía local (ríos, escorrentías superficiales, canales de riego, etc.) permite establecer un punto inicial en el estudio y diseño de planes de acción en lo referente a la protección del medio ambiente, recursos naturales y reingeniería de instalaciones de superficie. Se logra así un marco de referencia general para el tratamiento de los cruces de cañerías en un entorno geográfico formal, que además, permite instaurar programas con el objeto de evitar potenciales incidentes o acotar su impacto en caso de producirse.

INTRODUCCIÓN

La utilización generalizada de algoritmos genéticos de convergencia rápida para procesos de incertidumbre acotada en tareas que involucren la optimización de un modelo (en aplicaciones que van desde el proceso industrial de extracción y tratamiento de hidrocarburos al procesamiento de imágenes) encuentra en el tiempo de respuesta, tiempo de convergencia del algoritmo, un problema de difícil solución que, debido en este caso particularmente, al tamaño y complejidad de las imágenes, puede transformar esta técnica en inútil.-

Esta técnica se basa en las leyes de la selección natural, donde se premia la sobrevivencia a los organismos mejor adaptados. Estos organismos, representados en la imagen multiespectral como las celdas reconocibles a través de su textura y disposición geométrica, están sujetos a procesos de sobrevivencia, cruce, mutación y extinción. Para el cruce se pondera a los individuos y se beneficia con una mayor probabilidad de cruce a los individuos mejor adaptados, a continuación y en forma aleatoria se selecciona el punto de cruce. Este último paso podría ser mejor dirigido si se contara con una función de distribución de probabilidades que indique cuales son las porciones que mejor adaptación presentan, en este caso limitado a la extensión geográfica y características inherentes a la zona de estudio. Así, estas tendrían una mayor probabilidad de pasar a la siguiente generación sin ser seccionadas.-

Otra etapa de interés es la mutación, que para este caso es la representación de los cambios de textura y reflectividad de las celdas al acercarse a una planicie aluvial o cauce. Normalmente realizada en forma aleatoria en alguna celda de una pequeña parte de la población. La obtención de individuos mutados puede basarse en una función de distribución de probabilidades que indique la probabilidad que la mutación en una celda determinada mejore la adaptación del individuo.-

Teniendo presente que el propósito es el de acelerar la obtención de un individuo “óptimo”, al incluir el uso de funciones de distribución de probabilidad, el mejor de éstas debe realizarse con técnicas especiales que garanticen mantener la composición poblacional, con un número bajo de muestras. Para esto se ha recurrido al uso del método de muestreo por hipercubo latino(4).

DESARROLLO

Generación de la Población Inicial

El primer paso es la generación de la población inicial. Esta etapa es fundamental para lograr una rápida convergencia. Un pobre estudio estadístico del comportamiento de los datos de entrada dará por resultado una gran cantidad de generaciones antes de alcanzar el óptimo. Cada caso tiene un comportamiento particular y no se puede hablar de generalidades. Por ello las consideraciones que a continuación se realizan son válidas solo para los casos estudiados bajo las restricciones impuestas por las características fisonómicas propias de las trazas de cañerías y el contexto ambiental del yacimiento Loma La Lata.-

En nuestro caso particular han sido de interés dos problemáticas específicas del reconocimiento de formas en el tratamiento de imágenes multiespectrales con alta resolución espacial. En la primera, el objeto a reconocer tiene características estándar como la conformación geométrica (conos truncos con una orientación entre los 30° y 79°), textura, alta reflectividad (6). Este es el caso de las picadas (sendas de 4 metros de ancho o menos anteriormente utilizadas en tareas de prospección geofísica), caminos que rodean al borde (coronamiento) de la traza y el desmonte que implica la acometida como puede verse en la figura 1. Así, la forma de una acometida sobre la planicie aluvial del río Neuquén o algún componente hidrográfico, su acceso y coronamiento de la traza tienen características que nos permiten identificarla visualmente utilizando sistemas de interpretación de imágenes. De donde se pudo restringir la varianza en los datos dando un conjunto finito de formas geométricas como población inicial para el algoritmo. Para este caso el uso de mutaciones conducía a la degeneración de la población, por lo tanto, no fue utilizada.

En la segunda problemática enfrentada, el objeto a reconocer no tiene características tan estándar, no todas las obras se realizan del mismo modo ni siquiera se utilizan los mismos principios particulares (paso sobre el nivel de cota, paso bajo superficie, nivelaciones, inclinaciones, etc.), y el trabajo se complica ya que la imagen tratada tiene bordes difusos en algunos casos e indefinidos en otros (las acometidas soterradas comienzan su disposición hacia el cauce en forma anticipada con un desmonte no siempre significativo y en otros casos en proceso de revegetación). Este es el caso de las planicies de inundación o cercanías de cultivos intensos cercanos a cauces principales o canales de riego y desalinización como pueden ver se las figuras 2, 3 y 4. A diferencia del caso anterior, nos encontramos en un estadio previo al reconocimiento. El problema era tratar de determinar la forma aproximada o la forma patrón en cada caso, discriminando los distintos diámetros de cañería y estableciendo normas de comportamiento aproximadamente semejantes para las acometidas de cada tipología definida. Para este caso la población inicial surge de un tratamiento de imágenes convencional, lo que daba un punto de partida razonable, aunque con gran varianza en los datos. Se crearon áreas de entrenamiento definidas por el personal de Mantenimiento e Ingeniería del Yacimiento Loma La Lata y luego se utilizó el algoritmo Maximum Likelihood disponible en la herramienta de procesamiento de imágenes satelitarias. Una forma de enmascarar esta varianza para reducir su influencia en el tamaño de la población inicial fue aumentar la probabilidad de ocurrencia de mutaciones en aquellas celdas de la imagen cuya transformación tuviera una alta probabilidad de beneficiar al individuo.

La varianza en los datos de la población inicial pone de manifiesto la dificultad que se tendrá en obtener un individuo “óptimo”. Dificultad que se manifestará en el tamaño de la población inicial y

en el número de iteraciones. A esta problemática se la puede encarar adecuadamente con la variación y la forma de aplicación del operador mutación.

Individuos a Cruzar

La selección de los individuos a cruzar se debe realizar teniendo en cuenta la composición poblacional y permitiendo que los individuos mejor adaptados primero puedan sobrevivir a la siguiente generación sin alteraciones y además tengan una mayor posibilidad de cruce en la generación actual.-

Entonces el primer paso para la selección de los individuos a cruzar es obtener una función de distribución que represente la presencia ponderada de los distintos tipos de individuos en la población. Ya que no es lo mismo la presencia de un individuo 'A' que se encuentra "cercano" al "óptimo" que el de un individuo 'B' que se encuentra "lejano". Deberíamos premiar al individuo 'A' permitiéndole una mayor cantidad de cruces y esto se refleja en un valor de probabilidad más alto en la función de distribución.-

Para respetar la composición poblacional minimizando la cantidad de muestras, con el propósito de acotar el número de individuos en la población siguiente, se utiliza el método de muestreo por hipercubo latino (4). Este método consiste en la estratificación del eje de ordenadas de la función acumulada de la distribución poblacional en sectores equiprobables y disjuntos, igual a la mitad de individuos a obtener para la siguiente generación. Este proceso se realiza dos veces obteniendo dos conjuntos de 'n' pares de individuos a ser cruzados. A continuación se procede a una mezcla aleatoria del segundo conjunto. Finalmente se controla el nivel de correlación que existe entre ambos conjuntos y si se cumple con el valor preestablecido, entonces ya se cuenta con los pares de individuos a cruzar que se corresponden uno a uno entre ambos conjuntos.-

El nivel de correlación exigido entre ambos conjuntos es fundamental para una correcta generación de individuos. Un nivel alto de correlación obligaría a que los actuales individuos permanecieran prácticamente inalterados en sucesivas generaciones. Por el contrario un nivel de correlación bajo implicaría que los "mejores" individuos tiene afinidad por cruzarse con los "peores" individuos, algo que iría en contra del fundamento de la selección natural. Particularmente en este trabajo se utilizaron valores superiores a 0,58.

Cruce de Individuos

Normalmente el punto de cruce entre dos individuos se elige en forma aleatoria. Esta no es la forma más eficiente si se puede contar con una función de distribución que represente los puntos de cruce más favorables. El disponer de tal función no es trivial y de hecho debe ser recalculada en cada generación para cada par de individuos. Esto supone una alta carga de cálculo, que a primera vista tendería a hacer lento el proceso de convergencia. Para que esto no ocurra se deben utilizar en esta función parte de los cálculos realizados para la evaluación de cada individuo, obteniendo en ese momento, a la vez, el punto de cruce más beneficioso para cada individuo. Esta implementación no complica excesivamente la programación del algoritmo de cálculo y beneficia una rápida convergencia al "óptimo". El producto utilizado para realizar la programación del algoritmo fue el lenguaje denominado EML.

Al utilizar este esquema de cruce sobre las imágenes multiespectrales de alta resolución espacial se observó que el rendimiento del mismo era exponencial, alcanzando en pocas generaciones una alta velocidad de acercamiento al "óptimo". A continuación se pueden apreciar los resultados preliminares en las figuras 5 y 6:

Mutación de Individuos

La mutación se asume como la modificación aleatoria de una celda cualesquiera de la imagen. Esto permite tener acceso a nuevas soluciones que no podían obtenerse por el cruce de los individuos de

una determinada generación. Sin embargo permitir que tanto celdas “buenas” como “malas” sean afectadas por una mutación no es lo más conveniente. Más razonable es asociar una función de distribución que describa que tan “mala” es una celda y de allí seleccionar la celda a mutar. Esta función para el caso de obtención de acometidas y bordes de locación o caminos y coronamientos coincide con la utilizada para determinar en punto de cruce, lo cual implica que esta mejora no consume prácticamente mayor tiempo de procesamiento.-

Se utilizaron probabilidades de mutación desde 0,04 hasta 0,15. Según los resultados parece razonable utilizar hasta 0,10, pero esto es específico para el caso en estudio y las condiciones impuestas (5).

Es interesante notar que el hacer un uso adecuado de la cantidad de mutaciones por generación, puede permitir disminuir apreciablemente la cantidad de individuos de la población inicial. Esto a su vez facilita un acercamiento más rápido al “óptimo”. Se encontró que haciendo más propensos a la mutación las celdas “malas” de las “mejores” celdas se producía un ajuste “fino” en la obtención del “óptimo”.

Eliminación de Individuos

Durante este proceso los individuos cuya valoración supera a la media de la población, decrecen exponencialmente generación tras generación. Por lo que la eliminación de estos individuos en generaciones tempranas no afectará sensiblemente la evolución del resto de los individuos.-

La sentencia anterior no siempre es válida, pero para los casos analizados y la forma de implementación se cumple rigurosamente. En el caso de la localización para acometidas de cañerías, locaciones y caminos de acceso se cumple al trabajar con coordenadas planas absolutas, pero no así si se trabaja con coordenadas planas relativas y de manera inversa para el caso de los coronamientos.

CONCLUSIONES

Para la corroboración de los datos y la homologación de los resultados obtenidos se compararon las salidas de los procesos de información con los conocimientos de campo del personal de Producción e Ingeniería del Yacimiento Loma La Lata y verificaciones in situ utilizando Sistemas de Posicionamiento Global, tomando a estos últimos como casos resueltos.

En comparación con la implementación de algoritmos genéticos simples de convergencia lenta la mejora observada fue del 80% con un error del 0,125 % variando los tiempos entre 2 a 4,13 horas por cuadro de imagen y un rango de respuesta acertada del 87,5 % en la identificación de individuos (2), los resultados pueden verse reflejados en la figura 7 y 8:

Los resultados obtenidos alientan la utilización de este tipo de procesos en problemas con incertidumbre acotada, y especialmente en aquellos de convergencia lenta (por ejemplo en el inventariado automatizado de instalaciones de superficie relacionadas con las áreas de producción en hidrocarburos).

BIBLIOGRAFÍA

[1] Environmental System Research Institute Inc. (1990). Understanding GIS – The ArcInfo Method. p. 8-1 – 8-53.

[2] Stephen C. Guptill (1997). Elements of spatial data quality. p.167 – 186

[3] Paul Longley, Michael Batty (1996). Spatial Analysis: Modelling in a GIS Environment. p. 185 – 225

- [4] Douglas C. Montgomery, George C. Runger (2001). Probabilidad y Estadística aplicados a la Ingeniería. p. 500 – 675
- [5] Marc van Kreveld, Jürg Nievergelt, Thomas Roos, Meter Widmayer (1997). Algorithmic Foundations of Geographic Information Systems. p. 199 – 209
- [6] Emilio Chuvieco (2000). Fundamentos de Teledetección espacial. p. 165 – 204

FIGURAS



Figura 1: Ejemplo de disposición geométrica de un cruce sobre la margen del río

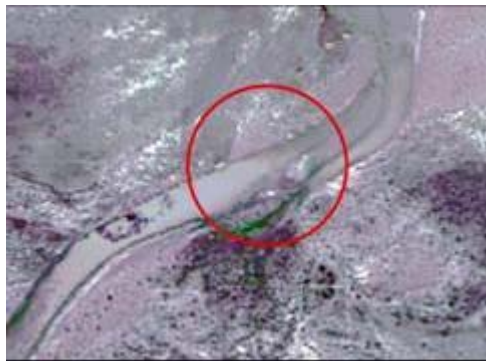


Figura 2: Ejemplo de instalaciones presentes sobre la margen del río



Figura 3: Ejemplo de instalaciones presentes sobre la margen del río

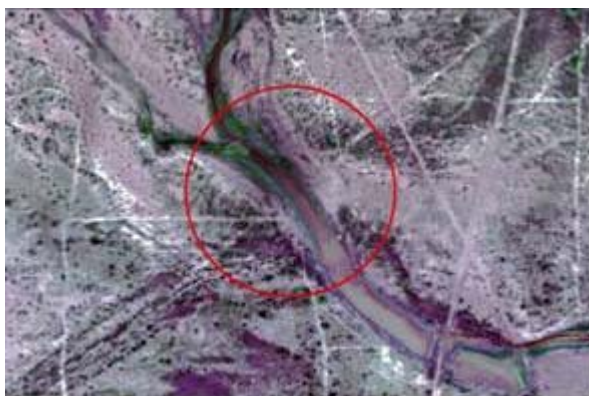


Figura 4: Ejemplo de instalaciones presentes sobre la margen del río

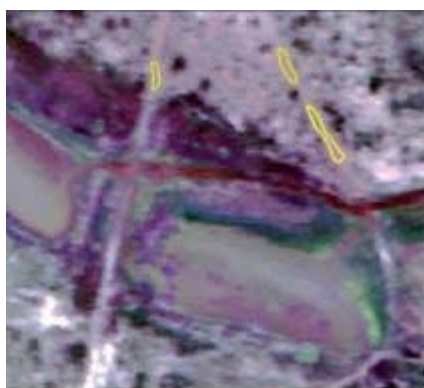


Figura 5: Ejemplo de resultados parcializados

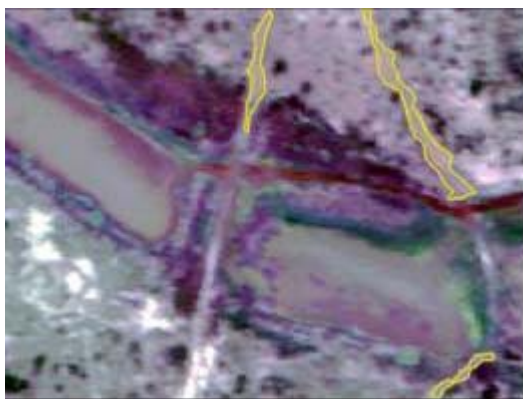


Figura 6: Ejemplo de resultados parcializados

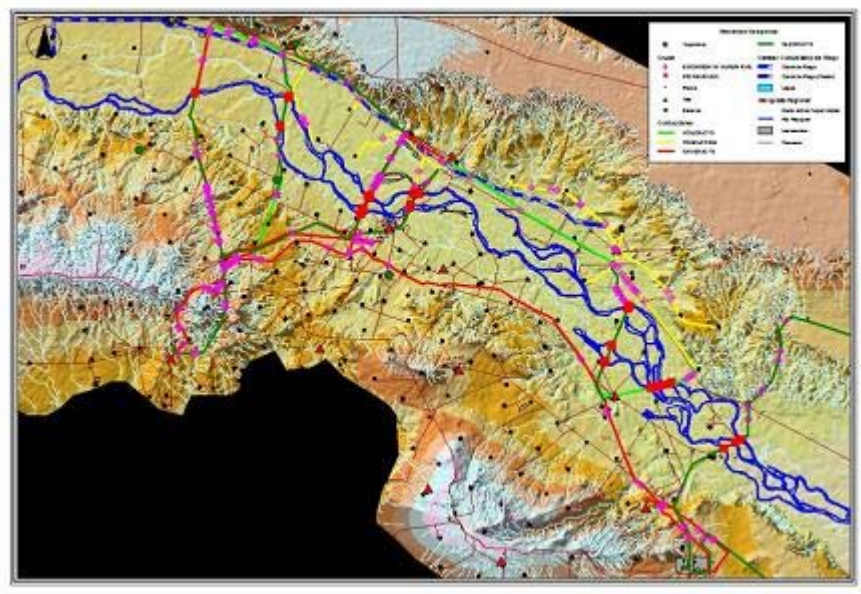


Figura 6: Cartografía 1:100.000 de los resultados obtenidos y modelo digital de elevaciones

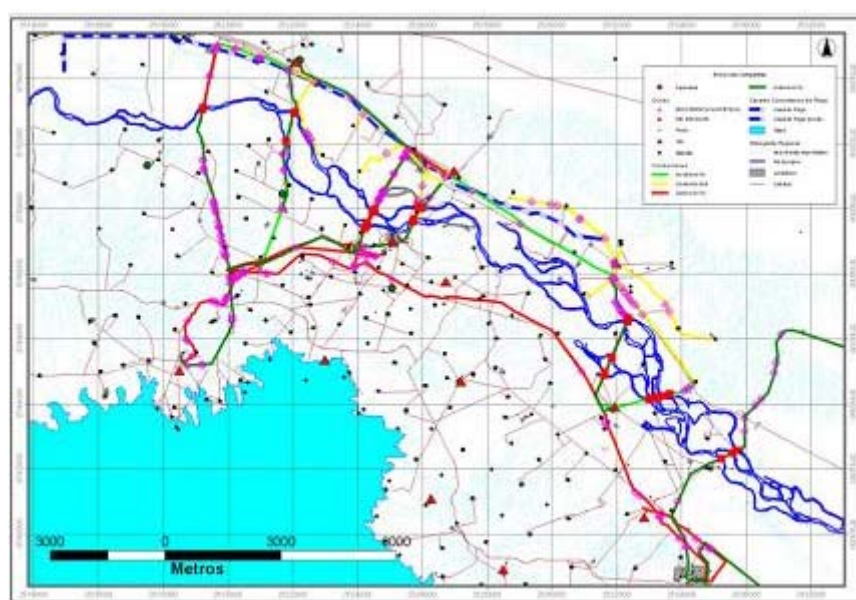


Figura 6: Cartografía 1:100.000 de los resultados obtenidos