

Diplomado en Análisis de Información Geoespacial

¿Cuáles son los componentes del dato espacial?

Autores:

M. en G. Daniel M. López López

M. en G. Aura C. Torres Gómez

Introducción

Los datos geospaciales son entidades espacio-temporales que cuantifican la distribución, estado y los vínculos de los fenómenos u objetos naturales y sociales, y se caracterizan por tener:

- Posición absoluta: sobre un sistema de coordenadas (x, y, z).
- Posición relativa: frente a otros elementos del paisaje (topología¹, incluido, adyacente, cruzado, se encuentra al oeste, norte, etc.).
- Figura geométrica que lo representa: punto, línea, polígono.
- Atributos que lo describen: características del elemento o fenómeno

Las características propias de los datos geospaciales dotan a estos de una gran potencialidad de análisis, al tiempo que condicionan o limitan otras operaciones.

Calidad de los datos

Trabajar con datos geospaciales tiene una serie de implicaciones que han de considerarse con detenimiento antes de llevar a cabo cualquier análisis.

Para aplicar con éxito los procesos de análisis espacial, la información georreferenciada debe tener calidad en términos de su accesibilidad, integridad, precisión, actualidad y consistencia, así como en sus fuentes de información y procesos de producción.

El desarrollo de métodos robustos para evaluar la precisión que permita validar los datos espaciales representa un reto para la comunidad de la ciencia geoespacial dada

¹ La topología expresa las relaciones entre los objetos de forma cualitativa: si dos polígonos son colindantes (contigüidad), si uno está contenido en el otro (inclusión), si dos líneas están conectadas (conectividad)" (Puebla y Gould, 1994).

nuestra capacidad actual para validar el gran conjunto de datos espaciales de la superficie terrestre representa un obstáculo para muchas aplicaciones futuras.

La preocupación por la calidad de los datos es básica por el simple hecho de que datos de mala calidad generan invariablemente resultados de mala calidad. Utilizar un dato de mala calidad es equivalente a utilizar un modelo equivocado.

Calidad, por lo general, es sinónimo de fiabilidad, pero en lo que respecta a los mapas significa que mantienen su integridad en relación con el mundo y se ajustan al uso que éste pretende de ellos. (Harvey, 2008, citado por IGAC-CIAF, 2002).

El procesamiento digital, principalmente para los Sistemas de Información Geográfica (SIG), ha puesto al descubierto las debilidades de los datos y la necesidad de crear normas y estándares para documentar su calidad, así como los productos que obtenemos a partir de ellos.

El error, que no es sino la discrepancia existente entre el valor real (puede ser un valor de posición, de un atributo, o cualquier otro), y el valor recogido en una capa. Los errores posicionales tienden a introducir anomalías en la combinación de múltiples conjuntos de datos o capas. La mayoría de las grandes superficies o coberturas de datos espaciales se derivan de los datos del sensor remoto y posteriormente son analizados en un SIG para proporcionar información de referencia que facilita el proceso de toma de decisiones (Lunetta and Lyon, 2004).

Según Kenneth y Huebner (1995), la calidad de los datos hace referencia a la relativa exactitud y precisión de una base de datos particular en un SIG.

La norma técnica colombiana 5043 (IGAC-CIAF, 2002) contempla la siguiente definición de calidad: “conjunto de características de los datos geospaciales que describen su capacidad para satisfacer necesidades establecidas e implícitas”.

La generación de información nueva, que puede provenir de un SIG, depende significativamente del contenido de la base de datos disponible. La calidad de esta base de datos determina la cantidad y calidad de los resultados del SIG.

Los usuarios de información espacial utilizan indistintamente los términos exactitud, precisión, resolución, detección y escala para referirse a la calidad o contenido de información de los datos. Esto es un error porque los conceptos definen propiedades o características diferentes de los datos.

- **Exactitud.** Es el grado en el cual la información de un mapa o una base de datos digital se muestra verdadera o con valores aceptables. La exactitud corresponde a la calidad de los datos y al número de errores contenidos en un conjunto de ellos o en un mapa. Analizando la base de datos de un SIG, es posible considerar la exactitud horizontal y vertical con respecto a la posición geográfica, tanto atributiva como conceptual, y a la agudeza lógica.
- **Precisión.** Hace referencia al nivel de detalle registrado o a la medida de las descripciones en las base de datos. Los atributos de información precisos pueden especificar las características de los elementos con gran detalle.

Es importante observar que los datos precisos, sin importar el cuidado de su medida, pueden ser inexactos, ya sea por un error humano o porque los datos pueden ser introducidos en las bases de datos incorrectamente.

Otro factor importante es la resolución que se define como el tamaño del rasgo más pequeño que se captura en los datos. En el formato raster esto es una función del tamaño de las celdas.

Por otra parte, el término compatibilidad indica que es posible usar dos o más conjuntos de datos. Mapas digitalizados a partir de fuentes de diferentes escalas pueden ser incompatibles. Para asegurar la compatibilidad, los conjuntos de datos se deben desarrollar usando los mismos métodos de captura, almacenamiento, manejo y edición.

El término aplicabilidad define la aptitud de un determinado conjunto de datos para un propósito particular.

Los datos de percepción remota están ampliamente disponibles, gracias a los avances en la tecnología satelital. Estos datos tienen diferentes niveles de exactitud y distintas resoluciones. El usuario debe decidir qué tipo de datos es apropiado para una aplicación particular.

Los resultados finales de la información extraída mediante la percepción remota, generalmente, se evalúan por su exactitud temática. En este sentido, el nivel de exactitud de los resultados se conoce y se documenta.

Según la norma técnica colombiana 5043 (IGAC-CIAF, 2002), los componentes para describir la calidad de un conjunto de datos son los elementos de calidad y los elementos generales de calidad.

Los elementos que se enlista a continuación son utilizados para describir los aspectos cuantitativos de calidad:

- **Totalidad.** Describe el nivel de veracidad de los elementos capturados, sus atributos y relaciones para representar el universo abstracto definido en las especificaciones de un producto.
- **Consistencia lógica.** Se trata del grado de certidumbre que permite a un conjunto de datos cumplir con las especificaciones de la estructura interna de los mismos y su topología.
- **Exactitud de posición.** Se refiere a la cercanía de los objetos en el conjunto de datos, con respecto a su posición verdadera (o la asumida como verdadera). La exactitud y precisión posicional son aplicables a la posición horizontal y vertical, y están dadas en función de la escala (vectores) o de la resolución (raster) en la que ha sido creado un mapa (impreso o digital).
- **Exactitud temporal.** Representa el grado de realidad en la escala de tiempo para los elementos de una base de datos con respecto a las especificaciones del producto.
- **Exactitud temática.** Se trata del grado de fidelidad que tienen los valores de los atributos asignados a los elementos de una base de datos con respecto a sus características en el mundo real, la clasificación correcta de los objetos y su relación con las especificaciones de un producto.

Jorge Fallas (Fallas, J. 2002) se refiere a este aspecto de la calidad como veracidad en la clasificación de los atributos. Los atributos son propiedades que poseen un determinado elemento geométrico del mapa (línea, punto o polígono). Estas variables se cuantifican a un nivel de medición nominal (clases de uso-cobertura del suelo) y de intervalo (grados de erosión) y, por lo tanto, el error no puede evaluarse en términos de la desviación estándar o de la raíz del error medio cuadrático (REMC).

Los sistemas de clasificación utilizados deben ser mutuamente excluyentes y contemplar la totalidad de las categorías presentes en el área de estudio. Cuando sea posible debe utilizarse un sistema de clasificación jerárquico que permita reagrupar

categorías con gran facilidad y mantener la lógica de la clasificación. Al reagrupar o generalizar clases se cumple con un estándar de clasificación preestablecido, por ello se deben utilizar indicadores naturales para definir las clases.

Para evaluar la exactitud de una clasificación (como en el caso de la cobertura vegetal y el uso del suelo, a partir de imágenes satelitales) es necesario comparar dicho mapa con información de referencia correcta; en estos casos se recomienda recurrir a una fuente de información más confiable que el material a evaluar.

El propósito de una evaluación cuantitativa de la exactitud es identificar y medir los errores de los mapas con base en un análisis coherente y objetivo.

La matriz de error es el formato más ampliamente aceptado para la presentación de informes sobre la exactitud de la clasificación de los datos de percepción remota.

Una matriz de error consiste en un conjunto de números, dispuestos en filas y columnas, que expresa la cantidad de píxeles o polígonos asignados a una categoría particular de una clasificación y en relación con los asignados a una categoría de otra clasificación.

Esta matriz muestra en su diagonal las categorías clasificadas correctamente; en las columnas se expresan las categorías deducidas por el autor de la clasificación y en las filas, las clases de referencia (verdad de terreno).

El error asociado a cada columna se denomina error de inclusión o comisión y representa las áreas que se asignaron a una categoría a la cual no pertenecían.

Por otra parte, en las filas se presentan las categorías verdaderas; el error asociado a cada fila se denomina de exclusión u omisión y representa las áreas que no fueron

asignadas a dicha categoría aun cuando pertenecían a ella. Este error le indica al usuario lo que puede esperar al utilizar el mapa en el campo.

- **Factores a considerar en la evaluación de la exactitud de la clasificación**

La matriz de confusión o el índice K permiten evaluar la exactitud de la clasificación, pero no brindan información sobre la calidad y veracidad de la información utilizada en el proceso.

A continuación se enuncian cinco aspectos que deben considerarse al evaluar la exactitud de la clasificación (Congalton, 1991):

- **Muestreo de campo**

Los datos de campo son considerados el estándar o valor real con el cual se compara la clasificación. De esta afirmación se desprende que si los datos no son de suficiente calidad y resolución, la evaluación será sesgada.

En algunos casos también es posible utilizar mapas previamente elaborados como material de referencia, en este caso la evaluación también dependerá de la veracidad del autor del mapa. Cada evaluación de exactitud es particular y, por lo tanto, para asegurar su éxito el investigador debe conocer su área de trabajo y determinar, basado en las limitaciones o facilidades de que disponga, la técnica de trabajo que le permita obtener la información de más alta calidad, al menor costo y en los plazos requeridos por el proyecto. El detalle con que se colecta la información de campo dependerá del detalle de la clasificación que se desee verificar.

- **Sistema de clasificación**

El sistema de clasificación utilizado para subdividir el área de estudio en unidades de menor tamaño responde a los objetivos del estudio y puede partir de sistemas preestablecidos como el de Anderson (1972), o diseñarse para responder a las características del área de estudio. A continuación se presentan algunos criterios que deben utilizarse al diseñar o seleccionar una clave de clasificación:

- El sistema de clasificación debe ser mutuamente excluyente y comprender la totalidad de las categorías de uso y cobertura presentes en el área de estudio.
- Cuando sea posible, se debe utilizar un sistema de clasificación jerárquico.
- El sistema jerárquico permite reagrupar categorías con gran facilidad y a la vez mantener la lógica de la clasificación. Reagrupar o generalizar clases permite cumplir con un estándar de clasificación preestablecido.
- Es necesario utilizar indicadores naturales para distinguir entre las clases.
- Por ejemplo, en el terreno es fácil diferenciar el pasto con arbustos del pasto sin arbustos; sin embargo, es más difícil separar pasto con cobertura de arbustos entre 0 y 25 por ciento que pasto con cobertura de arbustos entre 26 y 50 por ciento.

Finalmente, se debe recordar que antes de iniciar el proceso de clasificación es necesario tener una clara definición de las categorías. Cuanto más compleja sea la clasificación, mayor será el esfuerzo requerido para su verificación en el campo y el nivel de error de la clasificación.

Los elementos generales de calidad describen los aspectos cualitativos de un conjunto de datos. Los siguientes puntos deben ser utilizados para referir la calidad no cuantitativa de un conjunto de datos:

- **Propósito.** Registra los motivos por los cuales se creó un conjunto de datos y el uso que se tiene previsto para ellos.

- **Uso.** Documenta las aplicaciones que el productor (u otros usuarios) da al conjunto de datos.

Existen dos componentes en la calidad de los datos que deben ser considerados (Gutiérrez y Gould, 1994):

- **La temporalidad.** El tiempo es un componente esencial de los datos geospaciales y por tanto afecta su calidad. En general, la información debe ser la más actual y estar referida al mismo tiempo (Aronoff, 1989, citado por IGAC-CIAF, 2002).
- **La integridad.** Trata los criterios que se deben fijar y aplicar de forma adecuada para garantizar que las bases de datos sean totalmente homogéneas. Un aspecto estrechamente relacionado con la calidad de los datos es el error que puede haber en los mismos, el cual acompaña tanto a su imprecisión como a su inexactitud.

Según Kenneth y Huebner (1995), hay cuatro tipos de errores que es necesario conocer: de exactitud y precisión posicional, de exactitud y precisión de los atributos, de exactitud y precisión conceptual y de exactitud y precisión lógica.

- **Exactitud y precisión de los atributos.** Los datos no espaciales y la localización pueden ser inexactos o imprecisos. Esta inexactitud puede ser consecuencia de distintos tipos de errores. Los datos no espaciales pueden variar mucho también en precisión. La información precisa que los atributos describan los fenómenos con gran detalle.
- **Exactitud y precisión conceptual.** Los SIG dependen particularmente de la abstracción y clasificación de los fenómenos del mundo real. Los usuarios

determinan qué cantidad de información debe usarse y cómo ésta debe ser clasificada en las categorías apropiadas.

En ocasiones, los usuarios pueden usar categorías inapropiadas o clasificar erróneamente la información. Aun empleando las categorías correctas, los datos pueden estar mal clasificados.

- **Fuentes de inexactitud e imprecisión**

Las fuentes de error que pueden afectar la calidad del conjunto de datos de un SIG son muchas. Esto, que resulta muy obvio, puede no ser tan fácil de discernir. Algunas de estas fuentes de error serán automáticamente identificadas por el mismo SIG, pero es responsabilidad del usuario anticiparse a su aparición.

En algunos casos particulares se puede necesitar de comprobaciones específicas del error, porque los propios SIG son capaces de inducir al usuario una falsa sensación de exactitud y precisión, sin garantizar la validez de los datos.

De acuerdo con Burrough (1990, citado por Goodchild, 2007), los usuarios de los SIG deben reconocer el error, así como, el grado de error tolerable y asumible del sistema. Este autor divide las fuentes de error en tres grandes categorías: errores obvios, errores resultantes de la variación natural de las medidas originales y errores surgidos en los procesamientos.

Como fuentes obvias de error están:

- Antigüedad de los datos
- Área de cobertura
- Escala del mapa

- Densidad de las observaciones
- Relevancia
- Formato
- Accesibilidad

Entre los errores resultantes de la variación natural de los datos originales están:

- Exactitud posicional
- Precisión en el contenido
- Fuentes de variación de datos

Como fuentes de variación de los datos originados durante los procesos, se mencionan:

- Errores numéricos
- Errores en los análisis topológicos
- Problemas de clasificación y generalización
- Digitalización y errores geocodificados

Componentes de los datos geoespaciales

Los datos geoespaciales tienen tres componentes que hacen referencia a su localización, atributos y al tiempo.

- **Componente espacial**

Está ligada a la posición dentro de un sistema de referencia establecido de las propiedades de los objetos y a las relaciones espaciales que existe entre ellos. Esta componente es la que hace que la información pueda calificarse como *geográfica*, ya que sin ella no se tiene una localización, y por tanto el marco geográfico no

existe. La componente espacial responde a la pregunta *¿dónde?* a la ubicación geográfica, las propiedades espaciales de los objetos y a las relaciones espaciales que existen entre ellos.

Las relaciones espaciales se pueden clasificar en topológicas (de tipo cualitativo) y geométricas (calculadas a partir de las coordenadas de los objetos).

El ser humano tiene distintas formas de conceptualizar el mundo real. Cuando se desplaza de un lugar a otro, utiliza relaciones topológicas (cerca, lejos, dentro de, al norte de, etc.) pues se orienta por referencias relativas con respecto a un lugar conocido. Las relaciones topológicas más importantes son:

- Adyacencia (entre polígonos)
- Contigüidad (entre línea y polígono)
- Pertenencia (arcos a polígonos)
- Conectividad (entre arcos, en redes)
- Inclusión (punto en polígono, línea en polígono, polígono en polígono)

- **Componente temática**

Está ligada a los atributos de los objetos con los que representamos el mundo real. Cada objeto puede registrar un determinado valor para sus atributos (variables), los cuales pueden presentar cierta regularidad en el espacio y en el tiempo, además pueden ser de distinto tipo y escala de medida (Gutiérrez y Gould, 1994). Los atributos se expresan como variables, que pueden ser:

- Continuas: Admiten cualquier valor en un rango.
- Discretas: Sólo admiten valores en números enteros.

- Fundamentales: Se obtienen directamente del proceso de medición. Por ejemplo, el tamaño de una población.
- Derivadas: Se obtienen al relacionar dos o más variables fundamentales. Por ejemplo, la densidad de la población.

Para que las variables (atributos) puedan ser almacenadas en un Sistema de Información Geográfica (SIG), deben ser descritas mediante categorías. Esta descripción se realiza por medio de escalas.

- **Componente temporal**

La consideración de la dimensión temporal en un SIG supone la necesidad de almacenar y tratar grandes volúmenes de datos, ya que cada estrato, capa o nivel de información se debe almacenar tantas veces como momentos temporales se consideren para el análisis del área de estudio (Gutiérrez y Gould, 1994).

Referencias

Las notas fueron realizadas por el M. Daniel López y se realizó una adecuación.

- Anderson, James R., Hardy, Ernest E., and Roach, John T., 1972, A land-use classification system for use with remote-sensor data: U.S. Geol. Survey Circ. 671, 16 p., refs.
- Fallas, J. (2002) Normas y estándares para datos geoespaciales en Costa Rica. Recuperado el 5 de noviembre de 2014 de:
http://www.mapealo.com/costaricageodigital/documentos/alfabetizacion/normas_estandares.pdf
- Kenneth E. Foote and Donald J. Huebner. 1995. Error, Accuracy, and Precision. Department of Geography, University of Texas at Austin. Recuperado el 5 de noviembre de 2014 de:
http://www.colorado.edu/geography/gcraft/notes/error/error_f.html
- Madrid, Adriana y Ortiz, Lina María. (2005) Análisis y Síntesis en Cartografía: Algunos Procedimientos. Facultad de Ciencias Humanas. Universidad Nacional de Colombia sede Bogotá. 2005