

Determinación del tipo de proceso de remoción en masa en Bogotá mediante árboles de decisión

Determination of the type of landslide in Bogota based on decision trees

Alfonso Mariano Ramos Cañón¹, José Andrés Pacheco²

Resumen

Los modelos desarrollados a partir de herramientas de inteligencia artificial han sido utilizados de manera activa en los últimos años, siempre buscando capturar el conocimiento experto y/o información de mapas para reproducirlo en diferentes disciplinas y actividades. En este estudio se toma la información relacionada con la geología, geotecnia y mapas de pendientes, unidos con eventos de deslizamiento sucedidos en los últimos años en Bogotá para determinar un árbol de decisión. Dicho árbol de decisión, basado en la entropía de la información, entrega las bases para determinar zonas homogéneas en Bogotá, donde se prevé el desarrollo de un tipo o tipos específicos de proceso de remoción en masa. Además de la posibilidad de uso del procedimiento propuesto en otros ámbitos geográficos, el mapa presentado es una herramienta alternativa al estudio de la amenaza por deslizamiento en Bogotá.

Palabras clave: Árboles de decisión, deslizamiento, inteligencia artificial.

Abstract

Models developed through artificial intelligence tools has been used extensively in last years, always searching to capture expert knowledge and/or map information for use in other activities or disciplines. In this research we handle information relative to geology, geotechnique and slope maps, and also reports of landslide events for the past years in Bogota to determine a decision tree. The decision tree, based on entropy (information theory), is a useful tool to determine homogeneous zones in Bogota, where certain type of landslide could be developed. In addition to use this procedure in other geographic zones, the map obtained is an alternative tool to analyze the landslide hazard in Bogota.

Keywords: Decision trees, gliding, artificial intelligence.

1 Profesor Asociado. Instituto Geofísico. Pontificia Universidad Javeriana. E-mail: a-ramos@javeriana.edu.co
2 Estudiante maestría en Hidrosistemas. Pontificia Universidad Javeriana. E-mail: pachecoj@javeriana.edu.co

1. Introducción

En las últimas décadas los modelos predictivos de tipo heurístico han sido ampliamente usados con el objeto de generar una respuesta alternativa a problemas relacionadas con la materialización de diferentes amenazas, por ejemplo amenaza sísmica (Sikder *et al.*, 2009, Bommer y Scherbaum, 2008) y amenaza por inundaciones (Tokar y Johnson, 1999). En la temática de amenaza por deslizamientos autores como Arora *et al.* (2003), Lee (2007) y Rampone y Valente (2012) han usado Redes Neuronales Artificiales para generar mapas de susceptibilidad a deslizamiento. Aksoy y Ercanoglu (2011), Bui *et al.* (2012) y Pradhan (2012) aprovecharon las potencialidades de la lógica difusa para generar reglas que apoyen modelos de predicción de deslizamientos aplicadas a análisis de imágenes satelitales. Oh y Pradhan (2010) combinaron las redes neuronales y la lógica difusa para generar mapas de susceptibilidad a deslizamientos con base en imágenes fotográficas y de uso de suelo. Uno de los modelos que también pueden ser clasificados como predictivos de tipo heurístico son los árboles de decisión. Estos modelos han sido usados para clasificar los factores que inducen la ocurrencia de deslizamientos (Chu *et al.*, 2009) y para generar instrumentos de tipo analítico que permiten determinar pronósticos de deslizamientos (Nefeslioglu *et al.*, 2010; Saito *et al.*, 2009; Wang y Niu, 2009; Yeon *et al.*, 2010).

En este estudio se utiliza dicha herramienta para determinar el tipo de proceso de remoción en masa que se puede presentar basado en el análisis de información provista por la entidad encargada de realizar la gestión del riesgo por amenazas naturales en Bogotá (FOPAE – Fondo de Prevención y Atención de Emergencias). Posterior a la ocurrencia de un deslizamiento, profesionales del FOPAE realizan una visita técnica para generar un reporte de tipo técnico del acontecimiento. El conocimiento experto de las personas encargadas de evaluar cada evento de deslizamiento sucedido en la ciudad se encapsula en dicho reporte. La información relevante puede ser explotada a través de un análisis basado en entropía de la información materializada en lo que se denomina árboles de decisión.

Con base en el análisis de los resultados, se pueden identificar zonas homogéneas y factores que favorecen la generación de determinados procesos de remoción en masa. El contenido del artículo es el siguiente: la sección dos muestra el fundamento teórico para la construcción de los árboles de decisión, la sección tres presenta el procedimiento para la construcción de la base de datos de los procesos de remoción en masa junto con las etiquetas y atributos usados en el estudio. Finalmente se presentan los resultados en la sección 4 y las observaciones finales del trabajo.

2. Materiales y métodos

a. Árboles de decisión

Es un sistema experto que ayuda a generar un modelo de tipo predictivo basado en una serie de reglas sencillas que clasifican un conjunto de atributos de una base de datos determinada. Esto ayuda a identificar acciones a realizar en función del valor de una o varias variables (atributos) (Quinlan, 1986).

La construcción de un árbol de decisión se realiza en varias etapas:

Primero se revisa la base de datos sobre la cual se va a construir el mismo, conocer cuáles son las variables de entrada y cuál o cuáles son las de salida. Después se clasifica cada variable bajo una serie de etiquetas que identifiquen los diferentes valores que puede tomar (ya sea definiendo rangos numéricos en el caso de tener datos de este tipo, o directamente mediante datos discretos). Luego se identifican las cantidades de cada variable de salida para obtener lo que se denomina la entropía del sistema (Ecuación 1).

$$E = \sum_i^n -p_i \log_n p_i \text{ (Ec1)}$$

Donde E es la entropía del sistema, p_i es la probabilidad de ocurrencia y n es el número de variables de salida. Posteriormente se realiza la búsqueda de la variable de entrada que mejor clasifica la información de la base de

datos, es decir, la que indica la característica más relevante en el análisis de información (nodo raíz). Para esto se calcula la entropía de cada etiqueta dentro de cada variable de entrada y después se ponderan (para calcular una entropía promedio de cada variable). Por último se observa la diferencia que existe entre este cálculo y la entropía general del sistema; el que de una mayor diferencia (ganancia de información) es considerado como el nodo raíz del árbol. Una vez seleccionado el nodo raíz, se comienza el análisis de las ramas, es decir, para cada etiqueta del atributo principal se calculan las entropías de los otros atributos, y de esta manera se obtiene la variable para la cual se obtiene la mayor ganancia de información. Esta acción se realiza de manera cíclica hasta obtener la clasificación completa dentro del árbol de todos los datos de la base inicialmente tomada.

b. Base de datos de procesos de remoción en masa

El paso inicial para realizar el árbol de decisión fue obtener los informes de eventos de deslizamiento realizados por FOPAE para extraer la información relevante (atributos). De un listado inicial de 5839 informes suministrado por la entidad que contiene los nombres, código y tipo de evento, se realizó un filtro inicial por el campo de «tipo de informe», más exactamente por la opción «Remoción en masa» que son los eventos a los que se dirige este análisis, obteniendo una reducción significativa (1178 informes). Posteriormente se entró al portal de la entidad para descargar estos informes y realizar una revisión detallada del contenido de los mismos y así definir cuales eventos se deben a procesos de lluvia precedente al deslizamiento (FOPAE, 2012). Una vez se realizó esta acción, quedaron 113 informes de deslizamiento y se definieron los atributos que se podían extraer de allí, ya sea directamente por información del estudio o al ubicar el deslizamiento en mapas específicos que contienen información relevante como geología, geotecnia y el mapa de pendientes. Cada uno de los eventos fue ubicado en un mapa adquirido en el Instituto Geográfico Agustín Codazzi (IGAC), en formato DXF (mapa cargado en Autocad®) en escala 1:30000. Posteriormente se procedió a superponer los mapas de geología (Ingeominas, 1993) y de zonificación geotécnica de la ciudad (Universidad de Los Andes, 1996) para complementar la información necesaria que no se pudo obtener directamente de los informes. Los mapas geológicos y de zonificación geotécnica fueron obtenidos vía web (Observatorio Ambiental de Bogotá, 2012).

Con base en la información mencionada, se procedió a clasificar los 113 informes. Se seleccionó como variable de salida el tipo de deslizamiento. Sin embargo, solo fue posible utilizar 52 eventos ya que las descripciones de los restantes no brindan datos adecuados para definir el tipo de deslizamiento que se presentó. Las tablas de atributos y las etiquetas usadas en el análisis se presentan en las Tablas 1 a 4.

Tabla 1. Etiquetas del atributo de entrada «Pendiente»

Pendiente (%)	Etiqueta
0-20	Baja
21-45	Media
46-100	Alta

Tabla 2. Etiquetas del atributo de entrada «Geología»

Geología	Etiqueta
Formación Guadalupe - Cretáceo	Kg
Coluvión	Q(dp)
Aluvión Reciente - Cuaternario	Qal
Formación Chía	Qch
Cuaternario	Qrtsa
Formación Sabana – Pleistoceno	Qsa
Formación Guaduas - Cretáceo-Terciario	TKg
Terciario	Tma
Formación Usme - Eoceno Superior	Tu

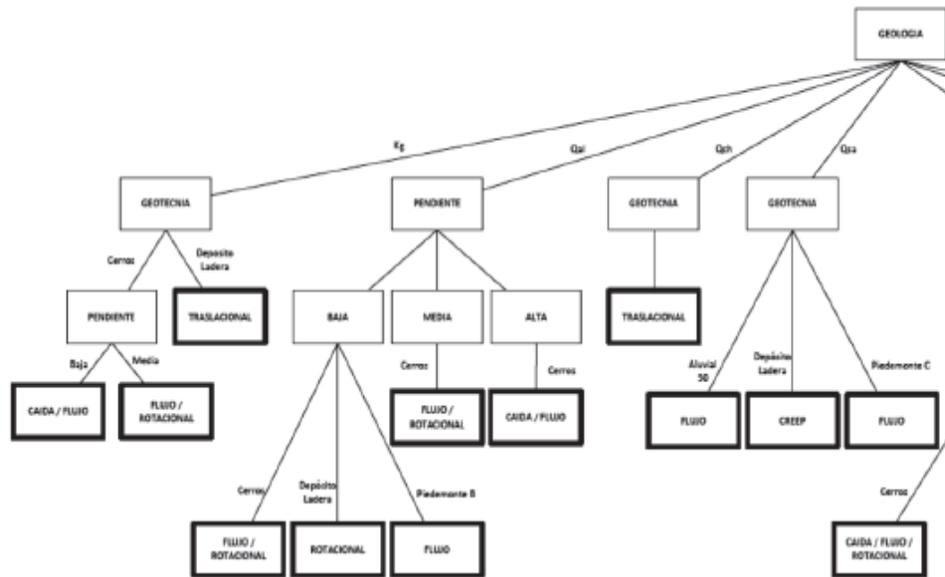


Figura 2. Acercamiento del lado izquierdo del árbol de decisión de la Figura 1.

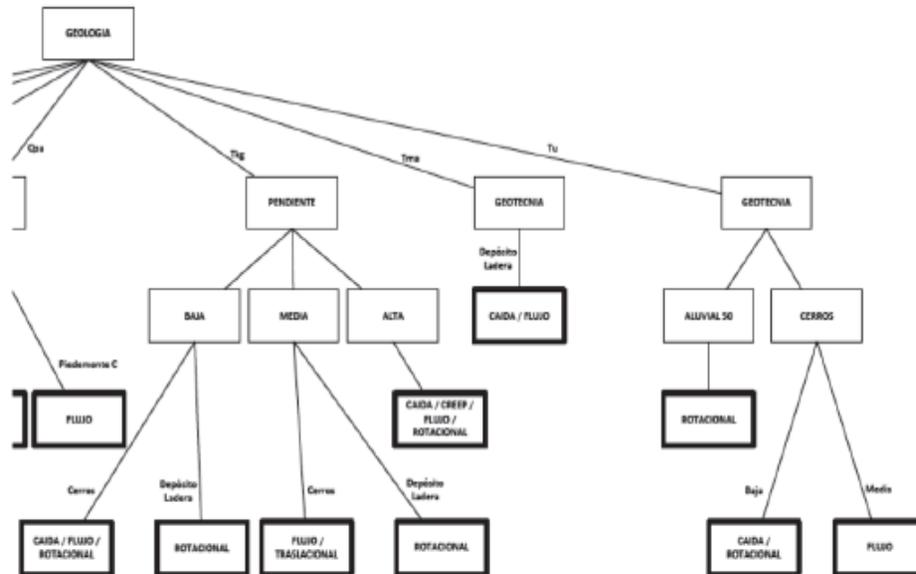


Figura 3. Acercamiento del lado derecho del árbol de decisión de la Figura 1.

El árbol de decisión mostrado en la Figura 1 es la herramienta usada para generar el mapa combinado de los atributos mencionados en las Tablas 1 a 3. El resultado del mapa es el tipo de deslizamiento que se espera en función del árbol de decisión encontrado. Adicionalmente, el uso del árbol de decisión en formato espacial permite detectar zonas en Bogotá en donde no se espera que se presenten deslizamientos dado que sus características de geología, geotecnia y pendiente no clasificaron. Es decir, la base de datos muestra que no se ha presentado deslizamientos bajo esta combinación de atributos en Bogotá.

La Figura 4 muestra el mapa resultado y un acercamiento en una zona determinada en el sur de Bogotá.

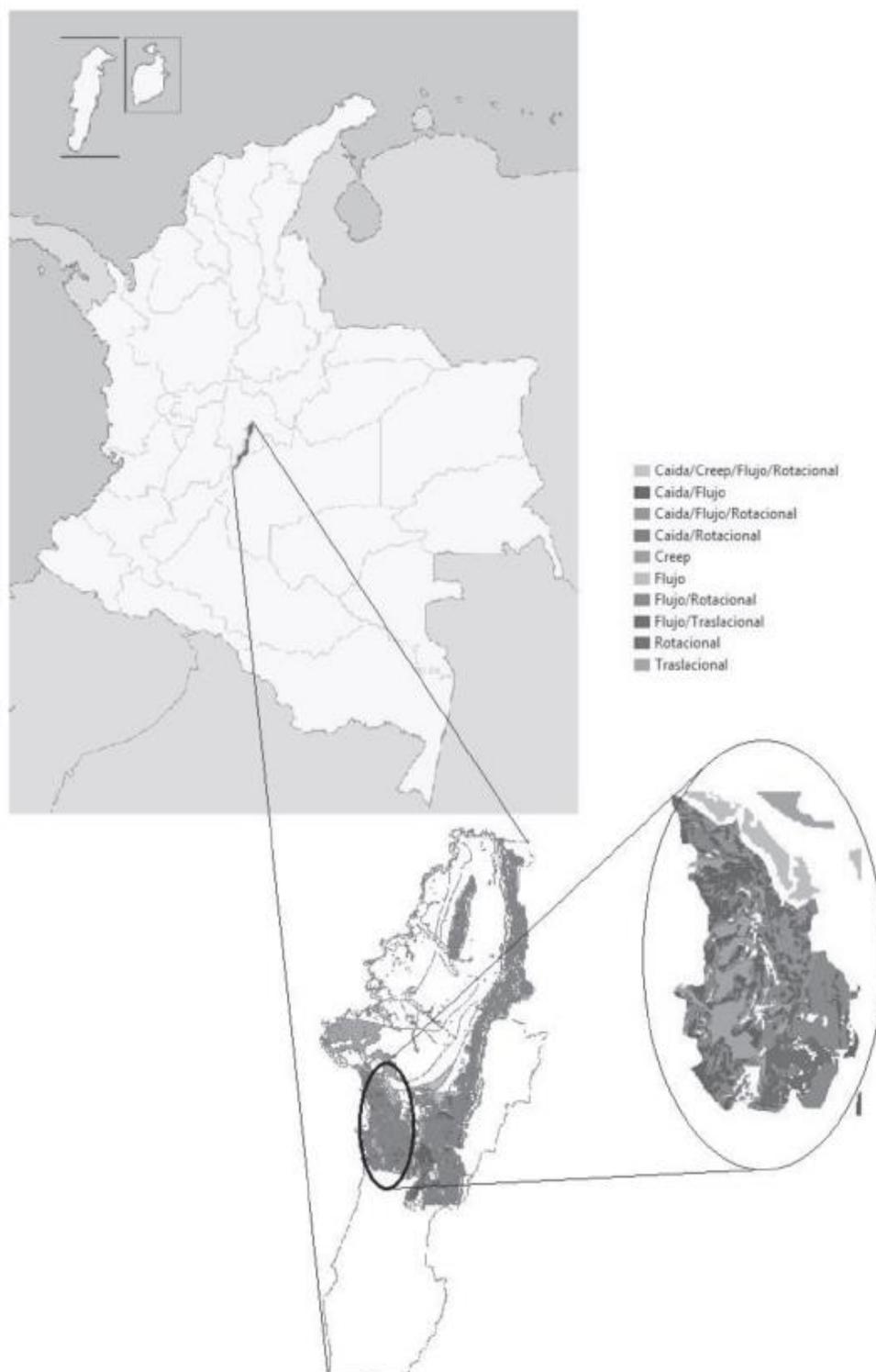


Figura 4. Mapa de tipos de proceso de remoción en masa para Bogotá.

4. Conclusiones

Al obtener un mapa que determina el tipo de proceso de remoción en masa que se presenta en Bogotá se pueden observar varios hechos:

Las zonas de los cerros son las que presentan una mayor susceptibilidad a este tipo de eventos, entre otras circunstancias debido al tipo de suelo que los compone (incluye escombros, rellenos de excavaciones y canteras anteriormente explotadas).

La predicción de este tipo de eventos es compleja y demanda una cantidad importante tanto de información como de variables no tenidas en cuenta en este estudio, que puedan orientar a una respuesta más precisa. Este ejercicio muestra la manera cómo podría usarse los árboles de decisión basada en la entropía de la información para generar zonas homogéneas de tipo de deslizamiento, así como determinación de zonas en donde las características de geología, geotecnia y pendiente no permiten que se presenten deslizamientos. Por lo anterior, el trabajo se muestra como una primera aproximación que refleja que ante la combinación de etiquetas en las variables presentes (i.e. Terciario Tma – Depósito Ladera – Pendiente Media) se pueden obtener diferentes tipos de deslizamiento (Caída o Flujo).

La generación de un árbol de decisión, al igual que un mapa de tipos de proceso de remoción en masa a partir de árboles de decisión son herramientas que eventualmente podrían usarse en la evaluación de amenaza por deslizamiento en cualquier zona. La explotación de bases de datos de deslizamientos con clara orientación de zonificación no es una tarea que se desarrolle comúnmente. Se propone un procedimiento que permite pasar del análisis puramente descriptivo de la base de datos de deslizamientos a un análisis que le posibilita al tomador de decisiones plantear alternativas en función de un análisis objetivo basado en entropía.

5. Agradecimientos

Los autores agradecen a la Pontificia Universidad Javeriana y al Banco Santander por el apoyo económico al presente trabajo de investigación a través del proyecto No 4756 «Retrospectiva de las catástrofes naturales en Colombia».

6. Referencias bibliográficas

1. Aksoy, B., Ercanoglu, M. 2011. Landslide identification and classification by object-based image analysis and fuzzy logic: An example from the Azdavay region (Kastamonu, Turkey). *Computers & Geosciences* 38, 87-98.
2. Arora, M.K., Das Gupta, A.S., Gupta, R.P. 2003. An artificial neural network approach for landslide hazard zonation in the Bhagirathi (Ganga) valley, Himalayas. *International Journal of Remote Sensing* Vol. 25, 559-572.
3. Bommer, J., Scherbaum, F. 2008. The use and misuse of logic trees in probabilistic seismic hazard analysis. *Earthquake Spectra*. Vol. 24 No 4 pp. 997-1009.
4. Bui, D.T., Pradhan, B., Lofman, O., Revhaug, I. 2012. Landslide susceptibility assessment in Vietnam using support vector machines, decision tree, and naive Bayes models. *Mathematical Problems in Engineering* Vol. 2012, 1-26.
5. Chu, C., Tsai, B., Chang, K. 2009. Integrating decision tree and spatial cluster analysis for landslide susceptibility zonation. *Engineering and Technology* 59, 479-483.
6. Fondo de Prevención y atención de Desastres (FOPAE), 2012. Informes del Sistema de Información para la Gestión de Riesgos y Atención de Emergencias de Bogotá (SIRE). Consultado el 5 de junio de 2012. <http://www.sire.gov.co/portal/page/portal/sire/>
7. Ingeominas. 1993. Mapa geológico. Proyecto de microzonificación sísmica de Bogotá. Convenio 01-93. Unidad para la prevención de emergencias del Distrito.
8. Lee, S. 2007. Landslide susceptibility mapping using an artificial neural network in the Gangneung area, Korea. *International Journal of Remote Sensing* Vol. 28, 4763-4783.
9. Nefeslioglu, H.A., Sezer, E., Gokceoglu, C., Bozkir, A.S., Duman, T.Y. 2010. Assessment of landslide susceptibility by decision trees in the metropolitan area of Istanbul, Turkey. *Mathematical Problems in Engineering* Vol. 2010, 1-15.

10. Observatorio Ambiental de Bogotá, 2012. Mapas de Geología y Topografía de Bogotá. Consultado el 15 de agosto de 2012. <http://oab.ambientebogota.gov.co/mig/map.phtml>
11. Oh, H., Pradhan, B. 2010. Application of a neuro-fuzzy model to landslide-susceptibility mapping for shallow landslides in a Tropical Hilly area. *Computers & Geosciences* 37, 1264-1276.
12. Pradhan, B. 2012. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Computers & Geosciences* 51, 350-365.
13. Quinlan J.R. 1986. Induction of decision trees. *Machine Learning* 1, 81-106.
14. Rampone, S., Valente, A. 2012. Neural network aided evaluation of landslide susceptibility in southern Italy. *International Journal of Modern Physics C* Vol 23, 1-20.
15. Saito, H., Nakayama, D., Matsuyama, H., 2009. Comparison of landslide susceptibility based on a decision-tree model and actual landslide occurrence: The Akaishi mountains, Japan. *Geomorphology* 109, 108-121.
16. Sikder, A., Iftikhar, U., Mumakata, T., 2009. Application of rough set and decision tree for characterization of premonitory factors of low seismic activity. *Expert Systems with Applications*. Vol 36 No 1 pp 102-110
17. Tokar, A., Johnson, P., 1999. Rainfall-runoff modeling using artificial neural networks. *Journal of Hydrologic Engineering*. Vol 4 No 3 pp 232-239.
18. Universidad de Los Andes. 1996. Mapa de zonificación geotécnica de Santafé de Bogotá. Proyecto de microzonificación sísmica de Bogotá.
19. Wang, X., Niu, R., 2009. Spatial forecast of landslides in three gorges based on spatial data mining. *Sensors* 9, 2035-2061.
20. Yeon, Y., Han, J., Ryu, K.H., 2010. Landslide susceptibility mapping in Injae, Korea using a decision tree. *Engineering Geology* 116, 274-283.