

Herramientas para la aplicación espacial de Árboles de Clasificación en el mapeo de la Cobertura de la Tierra

John Lowry, Christine Garrard, Douglas Ramsey and Vinod Chowdary
 RSGIS Laboratory, College of Natural Resources, Utah State University
 Logan, Utah 84322-5275

Resumen

El uso de árboles de clasificación (AC) para el mapeo de la cobertura de la tierra no es nuevo y de hecho se está haciendo más común. Los árboles de clasificación (a veces llamados árboles de decisión o CART (árboles de clasificación y regresión)) ofrecen varias ventajas sobre los algoritmos de clasificación tradicionalmente usados para el mapeo de la cobertura de la tierra. Una ventaja importante es la capacidad de utilizar con eficacia bases de datos pronosticadoras de orden categórico y continuo con diversas escalas de medida. Otras ventajas incluyen la capacidad de manejar datos de entrenamiento no paramétricos, buena eficacia de cómputo y una representación jerárquica intuitiva de las reglas de discriminación.

El algoritmo AC utiliza variables explicativas múltiples para predecir una sola variable de respuesta. En el caso de la cobertura de la tierra, las variables explicativas pueden incluir las bandas espectrales y sus transformaciones, datos derivados del DEM, y v datos vectoriales rasterizados, tales como geología o suelos. El algoritmo determina las características apropiadas de la variable de respuesta al fraccionar recurrentemente los datos explicativos en grupos cada vez más homogéneos (figura 1) con el resultado de producir un árbol jerárquico integrado por "reglas" (figura 2). El algoritmo AC es una herramienta que mina los datos y está disponible de varios vendedores comerciales, incluyendo: Insightful (S-PLUS™), RuleQuest Research (See5/C5.0®) y Salford Systems (CART®).

Un desafío importante al usar los árboles de clasificación para el mapeo de cobertura de la tierra es la aplicación espacial de las reglas generadas del software AC dentro de un sistema de información geográfico. A través del curso del proyecto de mapeo del uso de la cobertura de la tierra de Southwest Regional GAP, Utah State University desarrolló y/o evaluó tres herramientas para la aplicación espacial de las reglas AC.

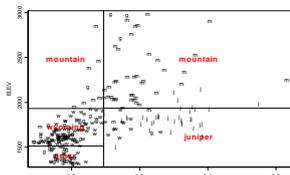


Figura 1. Fraccionar recurrentemente los datos con dos variables ('Elevation' y 'FALL 1999 NDVI').

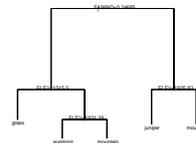


Figura 2. Ejemplo de árboles de clasificación de la Figura 1.

Este póster presenta brevemente estas herramientas. StatMod es una extensión de ArcView™ (ESRI) que tiene una interfase con S-PLUS para elaborar árboles de clasificación como rejillas (grids) de ArcInfo™. El módulo CART para ERDAS Imagine fue desarrollado por Earth Satellite Corp. para permitir la integración de la salida See5/C5.0 con ERDAS Imagine. Finalmente, RuleMaker es un libro basado en Internet que permite a un usuario traducir la salida del árbol de S-PLUS a un archivo *.mdl de ERDAS Imagine.

Referencias Seleccionadas

Breiman, L., J. H. Friedman, R. A. Olshen and C. J. Stone. 1984. *Classification and regression trees*. Belmont, CA: Wadsworth.

Brown de Colstoun, E. C., M. H. Story, C. Thomposon, K. Commisso, T. G. Smith and J.R. Irons. 2003. National park vegetation mapping using multitemporal Landsat 7 data and a decision tree classifier. *Remote Sens. Environ.* 85: 316-327.

Friedl, M. A. and C. E. Brodley. 1997. Decision tree classification of land cover from remotely sensed data. *Remote Sens. Environ.* 61: 399-409.

Hansen, M., R. Dubyah and R. DeFries. 1996. Classification trees: An alternative to traditional land cover classifiers. *Int. J. Remote Sensing* 17(5): 1075-1081.

Lawrence, R. L. and A. Wright. 2001. Rule-based classification systems using classification and regression tree (CART) analysis. *Photogrammetric Engineering & Remote Sensing*. 67(10): 1137-1142.

Pal, M. and P. M. Mather. 2003. An assessment of effectiveness of decision tree methods for land cover classification. *Remote Sens. Environ.* 86: 554-565.

Verbyla, D. L. 1987. Classification trees: a new discrimination tool. *Can. J. For. Res.* 17: 1150-1152.

StatMod Zone: Extensión de ArcView™ (ESRI)

Requisitos Del Software: Extensión de Análisis espacial de ArcView y S-PLUS.

Modo de uso: La extensión permite que usted someta trabajos automáticamente a S-PLUS. Esto significa que usted puede fijar todos los parámetros del árbol de clasificación, tales como poda y criterios de la validación cruzada, dentro de una caja de diálogo de ArcView. Estos se someten a S-PLUS donde se crea el árbol y entonces la salida de S-PLUS vuelve automáticamente a Arcview para la aplicación espacial. La salida final es una rejilla (grid) de ArcInfo™. El dataset de respuesta puede ser un shapefile de puntos o polígonos o una cobertura con un atributo que contiene el valor de la respuesta. Los datasets explicativos pueden ser shapefiles, rejillas (grid) de ArcInfo o coberturas.

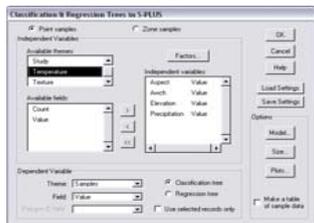


Figura 3. El caja de diálogo de ArcView donde se fija los parámetros del árbol de clasificación. En este ejemplo los variables (datasets) explicativos son: 'Aspect', 'Auch', 'Elevation' y 'Precipitation'. El variable dependent es el shapefile 'Samples.'

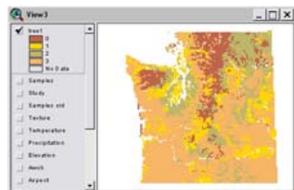


Figura 4. Ejemplo de resultados del árbol de clasificación.

Otras Características: StatMod Zone es capaz de generar árboles de clasificación y árboles de regresión usando S-PLUS. Si el usuario elige, hay capacidad de corregir las reglas del árbol antes aplicarlas espacialmente. Por defecto, un mapa de clasificación se produce, sin embargo, el usuario puede también producir un mapa de probabilidad o de "confianza". El mapa de confianza es una representación espacial de las probabilidades asociadas a cada celda siendo cartografiada correctamente. La extensión también se interconecta con el software estadístico SAS para la aplicación espacial de los modelos de regresión logística. **Autor y disponibilidad:** StatMod Zone fue desarrollado por Christine M. Garrard como cumplimiento parcial para un M.S. en Biología en Utah State University. La extensión de ArcView es gratis. La extensión y la guía del usuario pueden descargarse del siguiente URL: <http://www.gis.usu.edu/~chrisg/avev/>

El módulo Imagine CART:

Requisitos Del Software: ERDAS Imagine y See5/C5.

Modo de uso: El módulo fue desarrollado usando ERDAS Imagine EML y Lenguaje C de programación. El uso del modelo implica tres pasos. Primero, el usuario utiliza una caja de diálogo del módulo CART para crear la entrada en el software del árbol de clasificación. El conjunto de datos de respuesta es un archivo *.img de ERDAS Imagine con valores numéricos que corresponden a los atributos de respuesta. Los conjuntos de datos explicativos deben ser archivos *.img. Este paso produce un resumen de los datos en el formato apropiado de archivo para el See5/C5.0. El segundo paso implica usar el See5/C5.0 para generar el árbol de clasificación (Ej. El árbol de decisión). Una vez que se produzca el archivo See5/C5.0, el usuario regresa al módulo CART de ERDAS Imagine para aplicar espacialmente las reglas del árbol de clasificación generadas en See5/C5.0. La salida es un archivo *.img de ERDAS Imagine.

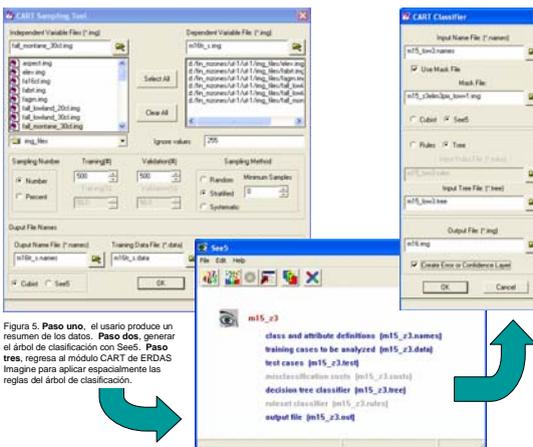


Figura 5. Paso uno, el usuario produce un resumen de los datos. Paso dos, generar el árbol de clasificación con See5. Paso tres, regresa al módulo CART de ERDAS Imagine para aplicar espacialmente las reglas del árbol de clasificación.

Otras Características: Al usar See5/C5.0 es posible utilizar el procedimiento llamado "boosting" que ha demostrado mejorar los resultados producidos de árboles de clasificación. Además de un mapa de clasificación, el usuario tiene la opción de producir un mapa de confianza, que es una representación espacial de las probabilidades asociadas a cada pixel que se dibuja correctamente.

Autor y disponibilidad: El módulo CART de Erdas Imagine fue desarrollado por Earth Satellite Corporation para el USGS (United States Geological Survey), EROS Data Center como ayuda para la base de datos nacional de cobertura de la tierra (NLCD) 2000. El software del módulo del CART para ERDAS Imagine se puede descargar gratis de: <http://www.gis.usu.edu/%7Eregap/download/C5Module/>

Libro basado en Internet RuleMaker:

Requisitos Del Software: ERDAS Imagine y S-PLUS.

Modo de uso: RuleMaker es un libro basado en Internet que se puede utilizar para traducir la salida del árbol de S-PLUS a un archivo *.mdl de Imagine. El usuario debe primero utilizar S-PLUS para crear el árbol de clasificación. El archivo de árbol debe ser guardado en el formato print.tree. El usuario va al Web site <http://www.gis.usu.edu/docs/projects/swgap/rulemaker.html>. RuleMaker requerirá que el usuario cargue la salida de árbol de S-PLUS, un archivo de variables, la dirección a donde se desea colocar el archivo *.mdl, y un nombre de la imagen de salida. Una vez que se corra RuleMaker (al presionar someter). El archivo *.mdl que resulta se descarga a su computadora y está listo para utilizarse en Imagine. La aplicación espacial de las reglas generadas desde RuleMaker se hace al correr el modeler en Imagine, produciendo un archivo *.img.

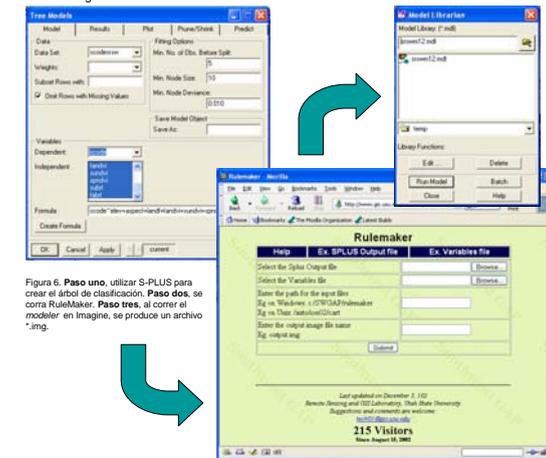


Figura 6. Paso uno, utilizar S-PLUS para crear el árbol de clasificación. Paso dos, se corra RuleMaker. Paso tres, al correr el modeler en Imagine, se produce un archivo *.img.

Otras Características: RuleMaker proporciona la opción de producir un mapa de confianza, que es una representación espacial de las probabilidades asociadas a cada pixel siendo dibujado correctamente. Dado que el archivo *.mdl de Imagine es un archivo de texto editable, las reglas se pueden editar por el usuario.

Autor y disponibilidad: RuleMaker fue desarrollado por Vinod Chowdary como cumplimiento parcial para un M.S. en Ciencias de la Computación en Utah State University. Usted puede tener acceso a RuleMaker en el Internet de manera gratuita en: <http://www.gis.usu.edu/docs/projects/swgap/rulemaker.html>