

ANÁLISIS DE LA INFORMACIÓN HISTÓRICA DE INCENDIOS FORESTALES PARA SU APLICACIÓN EN LA VALORACIÓN ECONÓMICA DE LOS INCENDIOS.

*Susana Martín Fernández, Luis G. García Montero, M^a Ángeles Grande Ortiz, Elena Blanco Martín, Sigfredo Ortuño Pérez, Pedro Hernando de la Cuerda.

ETSI de Montes

Ciudad Universitaria s/n

28040 Madrid

* susana.martin@upm.es

Resumen

El desarrollo de las tecnologías de la información y de las comunicaciones ha permitido la implementación de algoritmos eficientes para la extracción semi-automática de conocimiento de bases de datos y son el soporte de los sistemas de toma de decisiones. El objetivo de este estudio es analizar mediante técnicas cuantitativas las relaciones existentes entre las variables de las bases de datos de incendios forestales, y buscar e identificar información útil para la valoración económica del incendio y que facilite la posible tasación de un siniestro. Inicialmente se ha realizado una selección de las variables de interés cuantitativas y cualitativas para los objetivos planteados y que tuvieran suficiente información. Se han aplicado algoritmos de aprendizaje supervisado (análisis factorial, K-NN, K-means...) y no supervisado (análisis de la frecuencia, análisis de la dependencia, etc.), y análisis de series temporales, valorándose en cada caso la calidad de la información aportada por los resultados obtenidos con cada método. El estudio se ha aplicado a la base de datos de incendios forestales ocurridos en la Comunidad de Madrid desde 1968 hasta 2003 y han sido proporcionados por la Dirección General de Protección Ciudadana de la Comunidad de Madrid.

PALABRAS CLAVE: gestión del conocimiento, seguros de incendios forestales

MESA TEMÁTICA: 6

INTRODUCCIÓN

Las graves pérdidas económicas, ecológicas y sociales que provocan los incendios forestales, el incremento de la inversión en el medio rural, las políticas de transformación de terrenos agrarios en forestales, la fuerte demanda social de zonas recreativas, la implantación de los procesos de certificación forestal, entre otros factores, favorecen el desarrollo de un sistema de seguros contra incendios forestales por parte de las entidades aseguradoras de carácter privado.

El cálculo de primas de riesgo requiere conocer la frecuencia de la siniestralidad y el coste medio del siniestro (Picos, 2002). La frecuencia de la siniestralidad se obtiene a partir de los datos históricos de incendios forestales. La defensa contra incendios forestales es una actividad desarrollada o financiada por las Administraciones Públicas (Vélez, 2000). Las Administraciones han ido recopilando en sus bases de datos, la información procedente de los partes de incendios forestales. En estos partes aparecen la fecha del suceso, localización geográfica, tiempo de inicio, primer ataque y extinción, detección de las causas, condiciones meteorológicas, medios de extinción, superficie recorrida por el fuego, víctimas (Mérida, 2000).

El objetivo de esta comunicación es presentar los resultados de la implementación de algoritmos eficientes para la extracción sem.-automática de conocimiento de bases de datos para obtener conocimiento útil que permita hacer valoraciones económicas y nos acerque a la posible asignación de una tarifa a una zona rural a partir de información contenida en bases de datos de incendios forestales.

METODOLOGÍA

El aprendizaje automático se va a configurar a partir de los objetivos o intención del aprendizaje en sí mismo que aparecerá de forma implícita en el algoritmo de aprendizaje, del conocimiento previo del usuario ante ese problema concreto y de los datos u observaciones que se posean, su integración va a dar lugar a nuevo conocimiento. El problema de la extracción automática de conocimiento se resume en la figura 1 (Hernández, 2000):

Figura 1

La obtención de conocimiento a partir de bases de datos se nutre fundamentalmente de las siguientes disciplinas (Vidal y Hernández, 2000):

- ❑ Estadística
- ❑ Inteligencia artificial
- ❑ Visualización de datos
- ❑ Programación
- ❑ Interfaces de lenguaje natural a bases de datos

Las características de la base de datos de incendios forestales (número de variables cualitativas y cuantitativas muy alto, y conocimiento cronológico de los incendios), llevaron a la realización de los siguientes estudios:

- ❑ Análisis descriptivos para la caracterización de las variables
 - Análisis de la frecuencia de variables cualitativas
 - Análisis descriptivo de variables cuantitativas
- ❑ Preceptivos:
 - Predicción secuencial: Análisis de series temporales
 - Dependencia e interdependencia, relación entre las variables con el fin de seleccionar las que presentan una mayor interacción y estudiar la influencia de factores como causa del incendio, tipo de combustible,... en las variables de superficie quemada:
 - Estudio de la independencia de variables cualitativas, mediante el test Chi-cuadrado y el análisis del factor lambda.
 - Análisis de la varianza y test de rangos múltiples, Least Square Distances.
- ❑ Algoritmos de aprendizaje supervisado, con el fin de poder analizar agrupaciones de incendios forestales.

- o Análisis discriminante
- o K-means

RESULTADOS Y DISCUSIÓN

Para la realización de este trabajo se ha utilizado la base de datos de incendios forestales de la comunidad de Madrid, desde el año 1968 hasta el año 2003. Los datos del intervalo 1968-1973 no se estudiaron por falta de información. Las variables seleccionadas fueron las siguientes:

Variables Cualitativas: año, cuadrícula, causas, clase de día, orientación, especie, pendiente, modelo de combustible, lugar de inicio.

Variables cuantitativas: Superficie total quemada, Superficie total quemada monte alto (maderable), Superficie total quemada dehesas.

Este estudio incorpora variables no consideradas en otros estudios previos para la determinación de seguros contra incendios forestales. (Picos, 2002). El resto de variables no se han utilizado o bien por falta de información en ellas o bien porque no se consideraron necesarias para alcanzar a los objetivos planteados. Debido a lo exhaustivo del estudio realizado, se resumen a continuación los resultados más destacados:

Análisis de las variables cualitativas

Lugar de iniciación del fuego

El estudio de esta variable permite conocer en qué zonas es más probable el inicio de un incendio. La presencia o no de estas zonas en los terrenos a asegurar influirá en el cálculo de la tarifa.

Figura 2

Figura 3

La figura 2 muestra cómo el mayor porcentaje de incendios forestales se origina fundamentalmente en zonas próximas a carreteras y lugares del monte. Si reclasificamos los lugares de inicio del incendio, en lugares pertenecientes a la zona de interfaz urbano forestal (casas, urbanizaciones y basureros) y resto de zonas, vemos (figura 3) que hay una clara tendencia al aumento de los incendios en estas zonas. Existe una dependencia estadísticamente significativa entre la década en la que se han producido los incendios y la zona (Factor $\lambda = 0.017$).

Causa del incendio

Las causas de los incendios están en muchas ocasiones relacionadas con las actividades que se realizan en los terrenos forestales. Conocer la frecuencia de cada causa y las actividades desarrolladas en la zona a asegurar puede influir en la determinación de la tarifa.

Figura 4

Las causas de los incendios también han sufrido una evolución estadísticamente significativa ($p\text{-valor} = 0$ en el test Chi-cuadrado de independencia), a lo largo de los años, como se puede ver en la figura 4. Cada vez se identifican mejor las causas de los incendios y han ido aumentando causas como las debidas a las quemas agrícolas (21) y a los provocados, sin embargo, han disminuido los debidos a fumadores.

Relación causa del incendio con lugar de inicio

Existe una relación estadísticamente significativa ($p\text{-valor} = 0$ en el test Chi-cuadrado de independencia). En el caso de la zona de interfaz urbano forestal, las causas principales de los fuegos son otras negligencias, escapes de vertedero y quema de basuras, mientras que en el resto de zonas son las hogueras y los rayos los principales causantes.

Tipo de día de inicio de incendio

El tipo de día de inicio de incendio (festivo, víspera de festivo o laborable) no presentó ninguna relación estadísticamente significativa con el resto de variables.

Análisis de variables cuantitativas

Desde 1974 hasta 2003 ha habido 4988 incendios forestales en la CAM, con una superficie media quemada de 10,3 ha. y una desviación típica de 74.3 ha. Siendo los incendios más graves en 1985 cuando se quemaron 2783 ha. en el municipio de Torrelaguna y 2220 ha. en Robledo de Chavela.

Con el fin de detectar qué factores influyen en el tamaño del incendio, se ha analizado la influencia de las variables año, zona de inicio de fuego, orientación del terreno, causa del incendio, especie forestal, pendiente en las variables superficie total quemada, superficie total quemada de monte maderable y superficie total quemada en dehesas. Los resultados fueron los siguientes:

Superficie total quemada

- Análisis de la influencia de los siguientes factores mediante el test no paramétrico de Kruskal-Wallis:
 - No influyentes:
 - Década del incendio, (p-value: 0.003)
 - Tipo de zona de inicio del fuego (p-value: 0.06)
 - Orientación del terreno (p-value:0.6)
 - Influyentes:
 - Causas del incendio, (p-value: 0) Especie, (p-value:0)
 - Medio de detección, (p-value:0 .005)
 - Modelo de combustibles, (p-value: 0)
 - Tipo de pendiente, (p-value: 0)

Superficie Total Quemada de Monte Maderable

- Análisis de la influencia de los siguientes factores mediante el test no paramétrico de Kruskal-Wallis:
 - No influyentes:
 - Década del incendio, (p-value: 0.048)
 - Tipo de zona de inicio del fuego (p-value: 0.62)
 - Orientación del terreno (p-value:0.18)
 - Modelo de combustibles, (p-value: 0.056)
 - Influyentes:
 - Causas del incendio, (p-value: 0.004)
 - Especie, (p-value: 0.002)
 - Medio de detección, (p-value:0 .001)
 - Tipo de pendiente, (p-value: 0)

Superficie Total Quemada de Dehesas

- Análisis de la influencia de los siguientes factores mediante el test no paramétrico de Kruskal-Wallis:
 - No influyentes:
 - Década del incendio, (p-value: 0.19) , disminuyendo los últimos años la frecuencia pero aumentando la superficie quemada media del incendio.
 - Causas del incendio, (p-value: 0.48)
 - Especie, (p-value:0.06)
 - Medio de detección, (p-value:0 .42)
 - Tipo de zona de inicio del fuego (p-value: 0.62)
 - Orientación del terreno (p-value:0.57)
 - Tipo de pendiente, (p-value: 0.13), Baja frecuencia en zonas de gran pendiente, lo que provoca la homogeneidad del comportamiento medio de la variable
 - Modelo de combustibles, (p-value: 0.79)

Tanto en la superficie total quemada como la quemada en el monte maderable, influyen la causa del incendio, la especie, cómo se ha detectado y la pendiente de terreno, el modelo de

combustible influye claramente en la superficie total quemada. Se realizó un estudio más detallado de las Interrelaciones entre las variables, mediante el test de rangos múltiples (Martín, 2001). Los resultados mas relevantes fueron los siguientes: En el caso de superficie total quemada y el medio de detección del fuego, hay diferencias muy significativas entre la superficie media quemada cuando el medio de detección es 5 (34.5 ha) y cuando el medio de detección es 1 (15.8 ha). Para esta misma variable, pero considerando ahora los tipos de combustible, en los tipos 11, 10 y 15 la superficie media quemada supera las 70ha. En el caso del monte maderable, es para la especie 72, para la que la superficie quemada media los valores son muy elevados 214 ha.

Análisis temporal de la superficie quemada media anual.

Ningún coeficiente de autocorrelación ha sido estadísticamente significativo por tanto no se puede obtener un modelo predictivo.

Análisis discriminante

Con el fin de obtener una función discriminante para estimar la superficie quemada, se realizó un análisis discriminante, utilizando las variables más influyentes: lugar de inicio del incendio, causas, especie, pendiente, modelo de combustible. La variable superficie quemada se reclasificó según la tabla 1:

Tabla 1

La función discriminante obtenida fue la siguiente: $Y = 0,628008 * R52Bpendiente + 0,722875 * R53_MOD\ COMB$

Entre las 4387 observaciones utilizadas para ajustar el modelo, 3018 o 68,7942% fueron clasificadas correctamente. Este modelo sería orientativo respecto a qué superficie quemada se espera en función de las variables pendiente y combustible, el algoritmo de cálculo no ha considerado al resto de variables estadísticamente significativas.

Aplicados los algoritmos De K-means, no se obtuvieron resultados significativos que aportaran más información de los ya presentados.

CONCLUSIONES

El cálculo de primas de riesgos de incendios es un trabajo complejo en el que intervienen factores ligados al medio, administrativos y económicos. La riqueza de la información de las bases de datos históricas de incendios forestales, supone una fuente de conocimiento útil para la determinación de primas.

En este estudio se ha comprobado que en la asignación de tarifas a un terreno forestal es aconsejable considerar: la frecuencia de las causas de los incendios y la posibilidad de que estas causas se den en ese terreno, los medios existentes de detección, la localización de zonas con mayor probabilidad de que se inicie el incendio, las zonas de interfaz urbano-forestal que contenga, y por supuesto los modelos de combustible y la pendiente.

De los algoritmos aplicados, los más concluyentes, para el caso de la Comunidad de Madrid, han sido el test de independencia de chi-cuadrado, el test de Kruskal-Wallis, el test de rangos múltiples. No siendo aplicables o aportando menor conocimiento, el análisis de la varianza, el análisis de series temporales, el análisis discriminante y el método K-means.

Agradecimientos

Agradecemos a la Dirección General de Protección Ciudadana de la Comunidad de Madrid, a la Dirección General de Biodiversidad del Ministerio de Medio Ambiente y muy especialmente a D. Ricardo Vélez y a D. Juan Carlos Mérida, su siempre generosa colaboración en los estudios estadísticos sobre incendios forestales.

BIBLIOGRAFÍA

HERNÁNDEZ, J.; 2000. *Extracción automática de conocimiento en bases de datos e ingeniería*

del software. Universidad Politécnica de Valencia.

MARTÍN FERNÁNDEZ, S.; AYUGA TÉLLEZ, E.; GONZÁLEZ GARCÍA, C.; MARTÍN FERNÁNDEZ, A.; 2001. *Guía completa de Statgraphics*. Díaz de Santos, Madrid.

MÉRIDA, J.C.; 2000. Las bases de datos para definir el fenómeno de los incendios forestales. In: Vélez, R. (eds.) *La defensa contra incendios forestales: Fundamentos y experiencias*. 1.7-1.22. Mc Graw Hill. Madrid

PICOS, J.; 2002. *Los seguros de incendios forestales: Antecedentes y estudio de la viabilidad*. Tesis Doctoral de la Universidad Politécnica de Madrid.

VELEZ, R.; 2000. Perspectiva global: El fuego en los ecosistemas forestales. In: Vélez, R. (eds.) *La defensa contra incendios forestales: Fundamentos y experiencias*. 2.1-2.8. Mc Graw Hill. Madrid

VIDAL, E.; y HERNÁNDEZ, J.;, 2000. *Introducción a la prospección de datos masivos*. Universidad Politécnica de Valencia.

FIGURAS

Figura 1: Esquema del proceso de aprendizaje automático.

Figura 2: Diagrama de barras de la variable “Lugar de inicio del fuego”

Figura 3: Diagrama de barras de zona de inicio de incendios forestales (1 interfaz urbano forestal, 0 resto de zonas) y década.

Figura 4: Diagrama de barras de de las causas de incendios forestales a o largo de las 4 últimas décadas. (10- Rayo, 21 –quema agrícola, 22-quema de pastos, 23-trabajos forestales, 24-hogueras, 25-fumadores, 26-quema de basuras, 27-escapes de vertedero, 28-quema de matorral, 29-otras negligencias, 31-ferrocarril, 32-líneas eléctricas, 33-motores y máquinas, 34-maniobras militares, 35-otras, 40-intencionado, 50-causa desconocida, 60-incendio reproducido)

TABLAS

Tabla 1: Reclasificación de la variable “Superficie Quemada Total”.

Figura 1

Figura 2

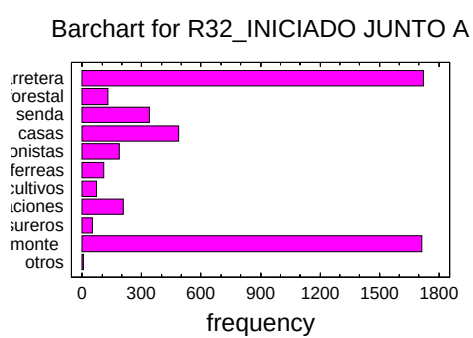


Figura 3

Diagrama de Barras para ANIO_1 según R32INICIADO_1

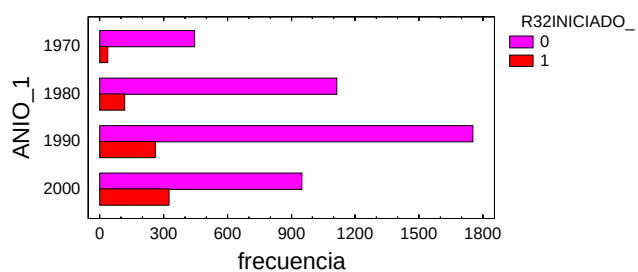


Figura 4

Diagrama de Barras para ANIO_1 según CAUSAS

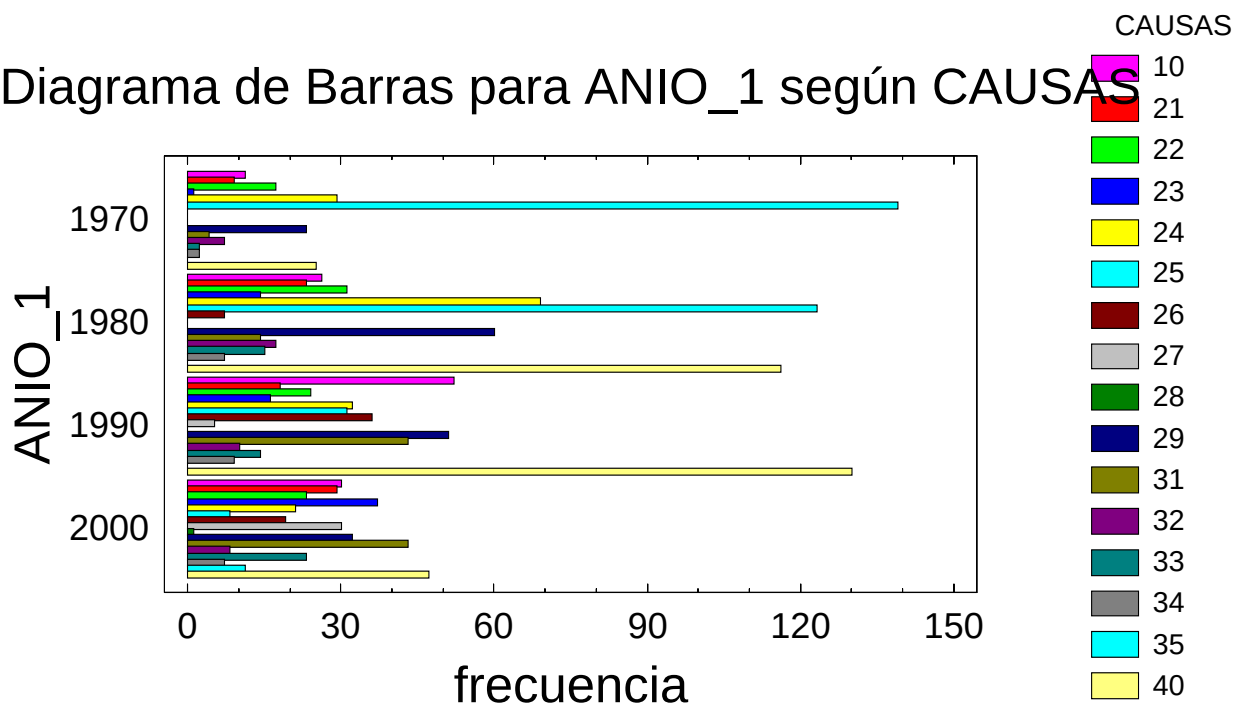


Tabla 1

SUPERFICIE (HA)	CÓDIGO
0-10	1
10-50	2
50-100	3
100-200	4
200-500	5
500-1000	6
>1000	7