



Universidad de Castilla-La Mancha

Escuela Superior de Informática

Departamento de Informática

Tesis Doctoral

**Contribución al estudio experimental de la predicción
basada en categorías deformables borrosas.**

Doctorando: José Angel Olivas Varela

Director: Dr. D. Enric Trillas Ruiz



Universidad de Castilla-La Mancha

Escuela Superior de Informática

Departamento de Informática

Tesis Doctoral

**Contribución al estudio experimental de la predicción
basada en categorías deformables borrosas.**

Junio de 2000

Doctorando: José Angel Olivas Varela

Director: Dr. D. Enric Trillas Ruiz

A mis padres,
que lo han dado todo por mí.

AGRADECIMIENTOS:

Mi más sincero agradecimiento a mi director, Enric Trillas, por su paciencia y dedicación, a Alejandro Sobrino, Claudio Moraga, Adolfo Rodríguez y Senén Barro por sus comentarios, a Ramiro Lago por las primeras implementaciones del sistema y a Francisco Pascual Romero por la ayuda prestada en la implementación definitiva.

También quiero agradecer a todos los miembros del Servicio de Defensa Contra Incendios Forestales de la Xunta de Galicia en Lugo (Técnicos, operadores, guardería, mantenimiento) su amistad y colaboración en el desarrollo de este sistema.

Por último quiero agradecer a mis compañeros de la Universidad de Castilla-La Mancha, mis compañeros de ICAI - Universidad Pontificia Comillas, mis familiares y amigos, el apoyo que me han prestado para hacer posible el desarrollo de este trabajo.

Índice.

Capítulo 1: Introducción.....	1
1. Modelo del Sistema	3
2. Lógica Borrosa e Inteligencia Artificial.....	5
3. Modelo informático para desarrollar el sistema.....	9
4. Planteamiento y justificación del trabajo: Aportaciones.....	11
5. Investigación en la lucha contra los incendios forestales.....	16
6. Marco de la tesis	21
Capítulo 2: Categorías Prototípicas Deformables Borrosas.....	23
1. Categorización conceptual. La noción de prototipo de Zadeh. Categorías Prototípicas Borrosas.....	26
2. Categorías Prototípicas Deformables Borrosas	54
3. Evaluación de situaciones reales mediante razonamiento no-monótono	60
Capítulo 3: Viabilidad y Descripción del Sistema INCEND-IA.....	65
1. Viabilidad del sistema INCEND-IA	65
2. Descripción del sistema INCEND-IA.....	72
3. La aplicación	84
4. Mapa de la aplicación	99
Capítulo 4: Evaluación de Casos Reales.....	115
1. Adquisición de Conocimiento: Obtención de datos fiables	118
2. Matching Caso Real – Categorías Prototípicas Deformables Borrosas. Aplicación a la Predicción de incendios forestales	184
3. Redistribución de Recursos	209
Capítulo 5: Validación y Verificación del Sistema	223
1. Evaluación del Sistema	223
2. Verificación del Sistema.....	224
3. Validación del Sistema	225

Capítulo 6: Conclusiones.....	253
1. ¿Qué problemas teóricos se han encontrado?. ¿Cómo se han resuelto?.....	254
2. ¿Han encontrado útiles los usuarios los resultados del sistema?.....	255
3. ¿Provoca el sistema mejoras en la calidad de las decisiones?.....	255
4. ¿Qué aproximaciones han sido inefectivas en este problema?.....	256
5. ¿Qué métodos han sido usados para evaluar la efectividad del sistema?	256
Bibliografía	259
Apéndice 1: Metodología, Conceptos y Herramientas	277
1. Método de trabajo	277
2. Adquisición de conocimiento y Data Mining.....	280
3. Razonamiento No Monótono y marcos.....	294
4. Razonamiento aproximado en IA.....	299
5. Sistemas de Información Geográfica	300
Apéndice 2: Incendios Forestales	303
1. Las nuevas tecnologías.....	306
2. Programa FUEGO.....	316
3. LATUV	320
Apéndice 3: Otras Aplicaciones del Método.....	323
1. Reconocimiento de la aparición de incendios forestales usando sensores de estado sólido.....	323
2. INFEDEC 2.1: Sistema de Ayuda en la Decisión de Diagnóstico y Tratamiento de enfermedades infecciosas.....	325
3. Otras aplicaciones.....	329
Apéndice 4: Incendios Producidos en 1992	331

Capítulo 1

I ntroducción.

La Inteligencia Artificial se puede considerar actualmente como una ciencia con un alto contenido experimental [Simon, 95]. Muchos de los programas que se generan en este ámbito surgen de experimentos con problemas que hay que automatizar. En bastantes casos, hay métodos algorítmicos, robustos en su funcionamiento, que producen resultados aceptables con un costo de desarrollo razonable. Algunos de estos programas reproducen lo que anteriormente se consideraba el resultado de un conocimiento experto adquirido empíricamente.

Como parte importante de la Inteligencia Artificial, la Ingeniería del Conocimiento es la disciplina que trata el desarrollo de los denominados Sistemas Basados en el Conocimiento, que se pueden describir como un tipo particular de sistemas software que tratan problemas cuyo método de resolución es más heurístico que algorítmico.

Muchos de los problemas que aborda la Ingeniería del Conocimiento se centran en comparar nuevas situaciones reales con situaciones similares ya conocidas o con patrones generados a partir de un conocimiento experto adquirido empíricamente. La resolución de estos problemas, habitualmente consiste en asimilar el comportamiento del nuevo caso al de los patrones o situaciones conocidas. Pero existen problemas reales, por ejemplo los de fenómenos que se comportan de forma cíclica, en los que es prácticamente imposible que una situación acontecida se repita exactamente en una nueva situación real. Por ello, parece conveniente

tratar de adaptar los patrones de situaciones ya acontecidas a nuevos casos reales, para que estos sean descritos con la mayor verosimilitud posible.

Por ello, este trabajo pretende ser una contribución experimental para analizar empíricamente cómo una situación nueva puede ser representada deformando los patrones (prototipos) que de situaciones similares se poseen. Si además los problemas a abordar tienen implícita una carga asociada de imprecisión e incertidumbre, la Lógica Borrosa proporciona herramientas para gestionarlos adecuadamente. El sistema que se presenta en este trabajo, fue abordado en un principio asimilando las nuevas situaciones reales a los prototipos establecidos. Pero el estudio más profundo del problema determinó que el uso de prototipos deformables borrosos propiciaba mejores resultados.

Por ello, el objetivo general de este trabajo es doble:

- Por un lado, realizar un desarrollo experimental que contribuya al estudio de los sistemas de predicción, mediante un Sistema Basado en el Conocimiento para predecir las pautas de aparición de incendios forestales en una *comarca gallega*, estimando el número de siniestros esperados para los siguientes días, así como las necesidades en cuanto a especialistas, patrullas y medios mecánicos, para poder combatir estos sucesos con eficacia.

El pronóstico resultante, será usado para optimizar el uso de recursos disponibles en la lucha contra los incendios forestales, tanto en ubicación, como en prioridades, horarios, etcétera.

- Por otro, intentar dotar este desarrollo de un modelo teórico, basado en categorías deformables borrosas, capaz de representar y manipular la complejidad que conlleva la resolución de este problema, obteniendo resultados realmente útiles (para los técnicos que han de tomar las decisiones) y manejables (información dada de una forma lo suficientemente clara como para ser usada de forma sencilla y directa, comprensible para todos los posibles usuarios del sistema).

Generalizar este modelo para poder ser utilizado con éxito en otros problemas y dominios.

1. Modelo del sistema.

El modelo utilizado para diseñar este sistema, es decir, la tarea o conducta a simular, es la realizada por los responsables en la lucha contra los incendios forestales (en este caso en Galicia), Ingenieros Técnicos o Ingenieros de Montes. El modelo se basa en la observación de cómo los técnicos humanos toman decisiones en condiciones reales. Hay dos aspectos destacables en este proceso:

1. Cuando un técnico se hace el planteamiento de una nueva jornada de trabajo, tiene en cuenta (consciente o inconscientemente) factores como los siguientes:
 - Cómo ha sido la aparición y evolución de los fuegos el día anterior.
 - Cómo han evolucionado los incendios desde que se ha iniciado un periodo de fuegos.
 - En qué estación se está y cuáles son las características específicas de la misma (puede ser un verano lluvioso o un invierno seco).
 - Cuál ha sido la capacidad de reacción ante nuevas alarmas con los recursos disponibles en las jornadas anteriores.
 - Se plantea cuestiones del tipo: ¿El horario de los recursos ha sido el óptimo?, ¿Se ha llegado lo suficientemente rápido a las nuevas alarmas?, ¿Ha habido alguna situación “peculiar” (por ejemplo la aparición masiva de alarmas en un área pequeña a alguna hora específica)?, ¿Ha habido o habrá algún acontecimiento social (fiesta, verbena...) que pueda trastocar el comportamiento habitual en la aparición y evolución de los incendios?, viento, etc.
2. El técnico establece una previsión de cómo puede ser el día o días sucesivos en cuanto a la ocurrencia de siniestros forestales, en base a las apreciaciones mencionadas y adapta sus decisiones comparándola con situaciones anteriores que ha conocido. Para ello, utiliza una especie de “prototipo de comportamiento”, que habitualmente resulta muy cercano a lo que realmente ocurre posteriormente.

Teniendo en cuenta estas reflexiones, el modelo que se plantea para el diseño de este Sistema Basado en el Conocimiento trata de reflejar este *modus operandi* de los técnicos humanos.

Tras el análisis del modo habitual de operar y de las características del entorno en el que se realiza la toma de decisiones, se pueden extraer una serie de conclusiones útiles de cara al intento de dotar de marcos formales a aquellas realidades que se quieren gestionar para abordar el problema.

La toma de decisiones se fundamenta mayoritariamente en datos aportados de forma imprecisa y subjetiva, motivados por la experiencia. Existen datos suministrados vía radiocomunicaciones por el personal cercano al fuego, cuando se trata de información relativa al incendio en sí, presentados en lenguaje natural, con valoraciones lingüísticas sobre las observaciones. Otros datos se obtienen de mediciones imprecisas (bien sea por la falta de precisión en la localización geográfica o en el tiempo, o por la propia capacidad de las técnicas o aparatos de medición), como puede ser la información meteorológica, los índices de peligro, la ubicación de los focos y de los recursos, la efectividad de los medios de ataque, etc.

A esta imprecisión suele añadirse un alto grado de incertidumbre, bien sea por la ausencia o baja calidad de la comunicación, falta de visibilidad y vaguedad en las informaciones o la inherente a algunos de los términos imprecisos usados.

Esta mezcla de vaguedad o imprecisión e incertidumbre en las entradas para el proceso de toma de decisiones, provoca que el mecanismo de razonamiento y las propias respuestas de los técnicos responsables sean de la misma naturaleza que las entradas, aunque con la suficiente entidad como para convertirse en mandatos.

Por otra parte, el uso de los mencionados “prototipos de comportamiento”, conlleva una notable dificultad para describir formalmente, verbalizar o plasmar de alguna forma estos patrones que el técnico utiliza. Esto es debido a la complejidad y número de los factores involucrados en este proceso de previsión, así como con la “velocidad” que hay que actuar.

Por ello, el uso de la Lógica Borrosa como modelo formal para gestionar la vaguedad, imprecisión e incertidumbre del razonamiento, junto con la teoría de prototipos, parecen herramientas formales adecuadas para abordar este modelado.

2. Lógica Borrosa e Inteligencia Artificial

Desde su concepción a finales de los años 50, el propósito fundamental de la I.A. fue construir programas que simulasen tareas que hacemos habitualmente los seres humanos, como apretar un tornillo o aprender a probar un teorema. De entre las tareas a simular alcanzaron sin embargo un protagonismo especial aquellas relacionadas con la simulación de procesos cognitivos o mentales considerados como inteligentes. Estos procesos comprendían con gran frecuencia, y como una característica señalada, razonamientos de sentido común. El razonamiento de sentido común se convierte así en un ámbito estrechamente vinculado al desarrollo de la I.A., que tiene como objetivo no sólo estudiar la habilidad de adquirir y aplicar conocimiento ordinario, sino de percibir y manipular objetos del mundo real.

Dos características señaladas de la I.A. en su propósito de estudiar el conocimiento de sentido común son:

- Tomar al ser humano como fuente de estudio a partir de la cual proponer modelos formales o computacionales:
 - Desde un punto de vista comportamental, fijándose en aquellas conductas de los humanos que expresan o son consecuencia de funciones cognitivas mentales.
 - Desde un punto de vista constitucional, aislando los elementos biológicos que permiten las funciones mentales y estudiando las conexiones que se establecen entre ellos; esto es, tratando de comprender la estructura neuronal humana. Minsky y Papert cuestionaron de modo muy influyente, por ser dos de los iniciadores de la I.A., la relevancia de este punto de vista; en particular, para el problema del aprendizaje en las redes neuronales multicapa, problema que contribuyó a superar Werbos al proponer el algoritmo de retropropagación.

Desde un punto de vista comportamental, los modelos formales de procesos mentales han puesto un extraordinario acento en la manipulación simbólica. El hallazgo de C. E. Shannon de la isomorfía existente entre las funciones de conmutación de circuitos con dos estados y las funciones booleanas de dos valores de verdad, le confirió desde el principio un extraordinario protagonismo a la lógica binaria como la herramienta principal para construir sistemas inteligentes. Al mismo tiempo, debido a lo fantástico de esta equivalencia funcional entre lo que se es capaz de hacer con una lógica y con un circuito, se generó en la comunidad de I.A. un rechazo hacia otros tipos de razonamiento y lenguajes de programación que no encajaban en este paradigma:

- Hacia el razonamiento cuantitativo o de métodos numéricos, de gran éxito en otros campos científicos y de ingeniería, con los que se han hecho grandes aportaciones a la humanidad, como construir presas o dirigir cohetes al espacio.
- Hacia razonamientos que intentan superar la dicotomía si/no, que limita la capacidad de un sistema para capturar procesos cognitivos humanos sofisticados, que usan una medida más compleja de comprensión, razonamiento y decisión sobre los problemas que la que supone una escala binaria.
- Hacia lenguajes de programación que no fuesen lenguajes de manipulación simbólica, de tal forma que hubo un tiempo en que casi era obligado expresar los problemas de I.A. en las máquinas LISP de McCarthy.

Las actitudes de rechazo hacia los métodos no binarios alcanzaron a lógicas que, como la lógica borrosa (LB), utilizaban -en opinión de fundadores de la I.A. como Minsky o McCarthy- un *armamentarium* errado. En efecto, por citar dos características fundamentales de la LB, a la cuantificación borrosa se la trata como un cálculo numérico y se introduce el concepto de “variable lingüística” como aquello que permite tratar el significado lógico de proposiciones de significado vago de una manera no dicotómica.

Una de las contribuciones mayores de la I.A. ha sido enfatizar la importancia del conocimiento en las acciones humanas inteligentes. Esto ha conducido a que una

de las áreas más importantes de la I.A. sea la de la representación del conocimiento. Hay un número extraordinario de propuestas interesantes a este respecto, pero todas ellas ven en cierto modo mermado su potencial al tener que atender, a la hora de su implementación, a la lógica binaria. Cuando se trata de modelar el razonamiento de sentido común, la comunidad de I.A. se dio cuenta muy pronto de las limitaciones de la capacidad inferencial de la lógica clásica (con la demostración de la existencia de dominios de aplicación de la lógica, como la aritmética, en la que hay proposiciones indecidibles), pero no de las limitaciones de una perspectiva binaria de los problemas y de su solución. Para ayudar a superar estas deficiencias surgió una nueva línea de investigación: el razonamiento no monótono. La no monotonía se vincula al hecho, tremadamente corriente en el razonamiento ordinario, de que la aparición de nueva información puede variar la conclusión inferida. Sin embargo, a causa de su aversión a la cuantificación numérica y su fuerte vínculo con la bivalencia, los sistemas no monotónicos desarrollados han sido especialmente complejos y difíciles de usar en aplicaciones prácticas.

Existen muchos métodos contrastados de representación de conocimiento: reglas de producción, marcos (*frames*), redes semánticas, lógica, sistemas de mantenimiento de la verdad,... Todos ellos, en su propuesta original, usan la lógica bivaluada de base. Quizás el punto de convergencia más interesante de la I.A. con la LB, sea la habilidad de esta última de proponer extensiones de muchos sistemas clásicos de representación de conocimiento en I.A. que permiten recoger la flexibilidad y variabilidad del razonamiento humano, muestra de lo cual es, por ejemplo, la valoración de proposiciones en términos de restricciones flexibles.

Dos de las herramientas de la LB que han permitido representar de modo flexible el significado vago y propagar la vaguedad en las inferencias, han sido la teoría del razonamiento aproximado y la lógica posibilística. En el Apéndice 1 se describen algunas de sus características más señaladas en cuanto a su papel en la I.A.

Uno de los desarrollos más prometedores de la I.A. quizá hayan sido los Sistemas Expertos (S.S.E.E.). Los S.S.E.E. son programas informáticos que constan de una

colección de reglas que describen el conocimiento acerca del dominio al cual se pretende aplicar. Un S.E. también tiene una base de hechos y un motor de inferencia, además de interfaces de comunicación adecuados.

2.1. Prototipos Conceptuales y Prototipos Deformables.

Tradicionalmente, el concepto de prototipo ha tenido mucha utilidad en el proceso de categorización conceptual del medio, al permitir agrupar a objetos relativamente dispares entre sí en clases comunes. Ello sucede, por ejemplo, con animales que, aún teniendo propiedades esenciales o accidentales diferentes, como sucede con un murciélagos o un gorrión, no obstante son clasificados como pájaros. En este ejemplo, el gorrión es usualmente considerado como el prototipo de la clase “pájaro”, ya que se presupone que reúne todas las características primarias y buena parte de las secundarias que se atribuye a la clase.

Los psicólogos cognitivos han reparado en los beneficios de la noción de prototipo como mejor ejemplar de la clase y, ya que se habla de clase, han verbalizado frecuentemente el problema en términos de conjuntos. Así, indican que un prototipo es aquel ejemplar que tiene pertenencia absoluta a la categoría o clase de la que es prototipo, clasificándose otros ejemplares por similaridad con él. La similaridad justifica el decir que su pertenencia a la categoría es verdadera con algún tipo de grado, que se cuantifica usualmente en razón de las coincidencias y divergencias entre propiedades esenciales y accidentales del prototipo y el ejemplar comparado.

Esta noción de prototipo, caracterizada en [Adán, 91] como “ingenua”, ha sido contestada en el ámbito de la lógica borrosa. Parece discutible que la ubicación conceptual de un objeto se haga siempre respecto a un ente ideal. En el ámbito del sentido común, no parecen existir entes absolutamente ideales, sino objetos que reúnen características de todo tipo, aunque algunas en mayor medida que otras.

Así, por ejemplo, un automóvil “Mercedes” coincide con una furgoneta o con un tractor en que posee cuatro ruedas. En esta característica coincide con un ejemplar mediocre (la furgoneta) y malo (tractor), del concepto “automóvil de turismo”. Un automóvil “Mercedes” tiene ruedas de un determinado tamaño de radio. En esto

coincide con un ejemplar mediocre (la furgoneta), pero no con uno malo (el tractor). Por último, posee una altura determinada, más pequeña que la de los otros ejemplares. No coincide, por tanto, con ninguno de ellos en esta característica. El prototipo de “automóvil de turismo”, -coincidamos en el Mercedes, por ejemplo- encierra características que también son comunes a ejemplares mediocres y malos del mismo. Esta es la idea que esta detrás de la propuesta de Zadeh [Zadeh 82], y que le valió para formular el concepto de prototipo borroso, una idea quizás menos intuitiva, pero más racional de cómo los humanos manejamos este concepto.

En el ámbito de la lógica borrosa se han hecho otras aproximaciones al concepto de prototipo. Una de ellas es la de *prototipo deformable*, originariamente propuesta por Bremermann [Bremermann 76]. En este contexto, se usa una noción ingenua de prototipo y se considera como deformable cuando se puede hacer un emparejamiento entre él y un objeto, de forma que el prototipo se adapte al objeto dentro de unos límites preestablecidos.

La noción de *prototipo deformable* surge, fundamentalmente, en el ámbito del reconocimiento de patrones. En ese contexto se considera, por ejemplo, que los caracteres escritos a mano pueden ser vistos como imitaciones distorsionadas de prototipos ideales, sobre los cuales existe bastante consenso en cuanto a su representación. Así, un carácter cualquiera, puede ser básicamente considerado como compuesto de segmentos cuya posición se controla con una serie de parámetros. Dado un carácter, se puede deformar su prototipo manipulando los parámetros hasta que coincide con el carácter en la mayor medida de lo posible, aunque es previsible la presencia de unos restos que no puedan ser cubiertos por las variaciones del prototipo.

3. Modelo informático para desarrollar el sistema.

Para simular el modo de razonar de los técnicos (expertos) en la lucha contra los incendios forestales con un sistema informático, se desarrolla un Sistema Basado en el Conocimiento, que además de las técnicas usadas habitualmente en este tipo

de desarrollos, incorpora nuevas aplicaciones de los conceptos provenientes de la Teoría de prototipos y la Lógica Borrosa anteriormente introducidos.

3.1. *El proceso KDD para extraer conocimiento útil de grandes volúmenes de datos.*

Actualmente, nuestra habilidad para analizar y entender conjuntos grandes de datos es muy inferior a la capacidad de almacenarnos. Una nueva generación de técnicas y herramientas computacionales se hace necesaria para la extracción de conocimiento útil, debido a que generalmente se da un rápido crecimiento del volumen de datos. Estas técnicas y herramientas son el sujeto de un nuevo campo de investigación denominado *Descubrimiento de Conocimiento en Bases de Datos* (KDD: *Knowledge Discovery in Databases*) y *Data Mining* (minería de datos) [Fayyad, 96a].

Tradicionalmente la forma de convertir datos en conocimiento radica en un análisis manual y una posterior interpretación. Normalmente este proceso resulta lento, caro y altamente subjetivo. De hecho se vuelve impracticable en muchos dominios como por ejemplo volúmenes de datos que crecen exponencialmente. Las bases de datos se incrementan en tamaño de dos formas distintas: por un lado el número de registros u objetos y por otro el número de campos o atributos por objeto. Existen bases de datos que contienen del orden de 10^9 objetos como por ejemplo en las ciencias relacionadas con la astronomía. En diagnóstico médico el número de campos puede ser fácilmente del orden de 10^4 . Esto motiva que haya que trabajar con miles de registros, cada uno de ellos con cientos de campos. El auténtico valor de estos datos reside en la habilidad del usuario para extraer informes útiles al realizar análisis estadístico e inferencia, y explotar los datos para obtener metas operacionales o científicas. Cuando la escala de manipulación de datos, exploración e inferencia crece por encima de las capacidades humanas se busca la tecnología de los ordenadores para automatizar del proceso.

El problema de extracción de conocimiento de grandes bases de datos, conlleva muchos pasos partiendo de la manipulación de los datos hasta la inferencia estadística, búsqueda y razonamiento. En 1989 se celebró el primer congreso sobre KDD y *Data Mining*.

La tarea de encontrar patrones en conjuntos de datos es conocida por diferentes nombres, según las diferentes comunidades científicas, por ejemplo extracción de conocimiento, descubrimiento de información, arqueología de datos, procesamiento de patrones de datos, etcétera. El término Minería de Datos (*Data Mining*) lo usan principalmente los investigadores en bases de datos, estadísticos y recientemente las comunidades empresariales y de negocio.

En este trabajo, se alude al término KDD para representar todo el proceso de descubrimiento de conocimiento útil a partir de datos, siendo el *Data Mining* un paso más de este proceso (aplicación de algoritmos específicos para extraer modelos de los datos), aunque se contemplan otros pasos como la preparación, selección y limpieza de datos, incorporación de conocimiento experto previo apropiado e interpretación de los resultados. Por esta razón, el KDD toma y aporta teorías, algoritmos y métodos de campos como las bases de datos, aprendizaje automático, reconocimiento de patrones, estadística, Inteligencia Artificial y razonamiento aproximado y Adquisición de Conocimiento en Sistemas Expertos [Fayyad, 96b].

4. Planteamiento y justificación del trabajo: Aportaciones.

El planteamiento de este trabajo consiste en tratar de simular el modo de razonar de los técnicos del ámbito de la lucha contra los incendios forestales, utilizando para ello modificaciones y adaptaciones de las herramientas y técnicas brevemente descritas, en el marco de una metodología genérica de desarrollo de Sistemas Basados en el Conocimiento. Se diseña e implementa un sistema de ayuda a la decisión para la gestión de incendios forestales y se prueba en condiciones reales.

Aunque se verá con sumo detalle en el test de viabilidad de la aplicación, la relevancia social del trabajo, en la medida que pueda ayudar a reducir efectos ecológicos o de daños humanos en los siniestros forestales, está más que justificada. Además, el interés científico, pasa por la presentación de un nuevo marco para sistemas de predicción, basado en la modificación de prototipos hacia situaciones reales, que permitan establecer previsiones.

4.1. Hipótesis y objetivos.

La hipótesis de trabajo es que es posible mejorar los sistemas informáticos de ayuda en la lucha contra los incendios forestales, utilizando adecuadamente técnicas y herramientas de Ingeniería de Conocimiento, y en general técnicas de Computación avanzadas y de Inteligencia Artificial. Para ello, se adaptan o extienden (como medio, no como fin) los formalismos enumerados anteriormente (Ingeniería del Conocimiento, Lógica Borrosa, KDD, *Data Mining*, Prototipos, Prototipos Deformables) para tratar el problema específico que se quiere solucionar.

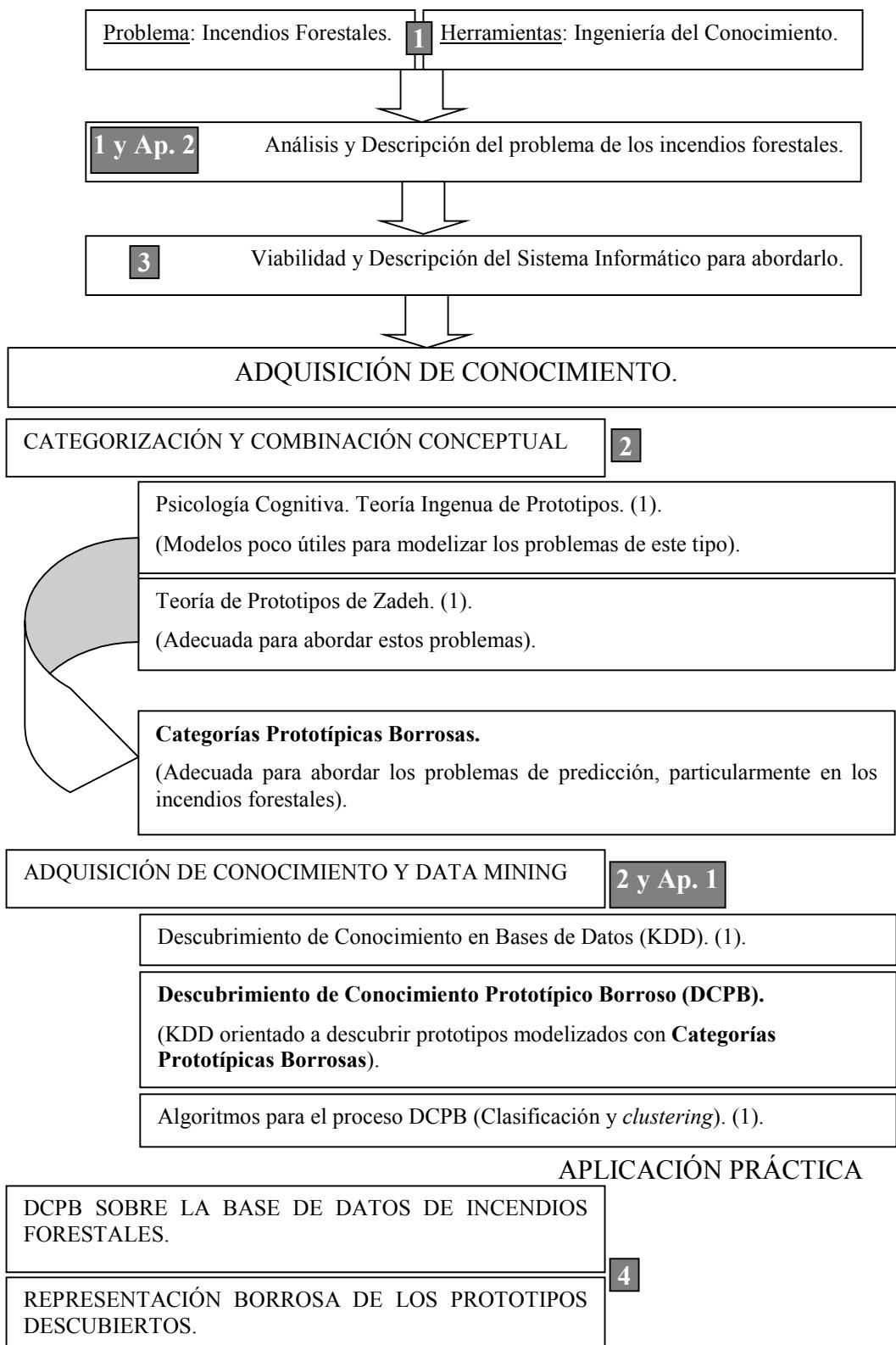
El objetivo principal de este trabajo es el diseño e implementación de un prototipo de Sistema Basado en el Conocimiento para la gestión y ayuda en la decisión en la lucha contra los incendios forestales, que mejore las prestaciones y la adecuación a la realidad de los sistemas informáticos que se utilizan habitualmente.

Este objetivo se puede articular en los siguientes pasos:

- Realizar un análisis en profundidad de la problemática en la lucha contra los incendios forestales, y en detalle de las tecnologías utilizadas.
- Proponer un nuevo mecanismo para abordar el problema de la predicción del riesgo de incendios y nuevas técnicas para utilizar de forma óptima los recursos de los que se dispone en la lucha contra los incendios forestales.
- Verificar y validar los métodos desarrollados mediante el diseño e implementación de un prototipo de un Sistema Basado en el Conocimiento, que gestiona con detalle una unidad geográfico-administrativa estándar (comarca, en Galicia existen 43 comarcas), y que puede ser probado en casos reales en los centros de coordinación gallegos.

4.2. Esquema global.

El esqueleto de la memoria que se presenta es el que describe la figura 1, donde los números indican el capítulo o apéndice donde es abordado el tema y se resaltan en negrita las aportaciones teóricas originales de este trabajo:



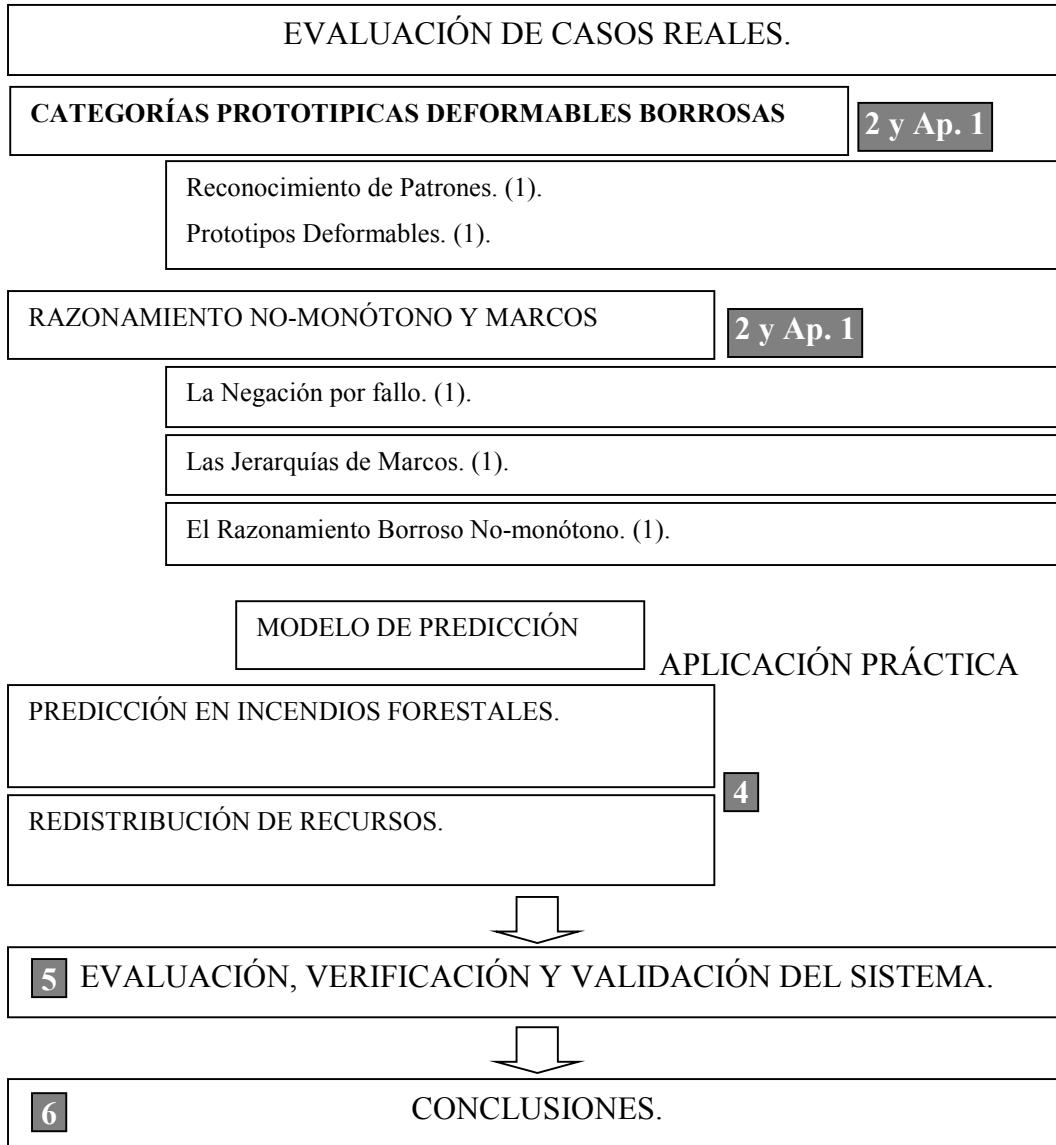


Figura 1: Esquema general del trabajo.

- (1) Se describen someramente con el fin de ser utilizados en la práctica. No se detallan ni analizan en profundidad.

4.3. Aportaciones de la tesis.

Se desarrolla un prototipo de Sistema Experto para la predicción y la optimización de recursos en la lucha contra los incendios forestales, cuyas principales aportaciones novedosas son las siguientes:

4.3.1. De fundamentos teóricos.

En el proceso de Adquisición de Conocimiento, se usan nuevas extensiones de técnicas clásicas como son Emparrillados (*Repertory Grids*) Borrosos y Cuestionarios especialmente diseñados para educir conocimiento prototípico y conocimiento impreciso.

Pero quizá lo más relevante en esta fase sea la introducción de un nuevo mecanismo que consistirá en modificaciones de algoritmos de aprendizaje clásicos, especialmente ID3 y CART, y creación de nuevos algoritmos especialmente diseñados para tal propósito, y se denominará **DCPB** (basado en KDD y *Data Mining*): **Descubrimiento de Conocimiento Prototípico Borroso**, que se aplicará a las grandes bases de datos sobre incendios forestales de las que se dispone.

A la hora de representar el conocimiento adquirido, se usa, entre otras, una base de conocimientos basada en razonamiento no-monótono y marcos, usados de dos formas diferentes: Como *colección* de marcos, cada uno de ellos correspondientes a un prototipo conceptual o como *jerarquía* de marcos, cada marco es o bien un prototipo, o una clase de ellos. La definición de estos prototipos se hace en base a la definición de lo que se denominarán **Categorías Prototípicas Borrosas** (basadas en la noción de prototipo borroso introducida por Zadeh en 1982). La modelización de las nuevas situaciones reales (que servirá para la predicción, ayuda a la decisión y optimización de recursos), está basada en las Categorías Prototípicas Borrosas y en un concepto que proviene del ámbito del reconocimiento de patrones, el concepto de prototipo deformable, que nunca ha sido usado para este menester. Se presentarán entonces las que serán denominadas **Categorías Prototípicas Deformables Borrosas**, que describirán las situaciones reales adaptándose a ellas.

4.3.2. De diseño del sistema.

El sistema esta soportado en una plataforma visual, basada principalmente en el lenguaje C++. Requiere un tratamiento específico la gestión de bases de datos de un tamaño considerable y el tratamiento gráfico, parte esencial del interfaz del

sistema. Esto provoca la necesidad de una implementación muy eficiente, con una estructuración muy cuidada. No es frecuente la implantación de estas novedades conceptuales en un sistema de este nivel de complejidad. La metodología utilizada para el desarrollo, aunque basada en una genérica (IDEAL, Apéndice 1), tiene modificaciones que permiten un desarrollo más coherente y con más garantías de éxito de Sistemas Basados en el Conocimiento, y ha sido probada en más de 10 sistemas reales similares (Apéndice 3).

1.1.1 De prestaciones del sistema.

El tratamiento de información social e imprecisa permite dar respuestas más precisas y más útiles a los usuarios del sistema. Esto es absolutamente novedoso en cuanto a los sistemas que existen para abordar la lucha contra los incendios forestales. Además, el modelo de predicción es exportable a otras muchas posibles aplicaciones, algunas ya desarrolladas (Apéndice 3).

5. Investigación en la lucha contra los incendios forestales.

La principal característica que distingue el problema de la lucha contra los incendios forestales es su complejidad. En la aparición de incendios forestales están implicados un gran número y variedad de factores, los naturales, los humanos y los sociológicos, y deben ser considerados simultáneamente para poder diseñar un modelo del problema lo suficientemente completo como para que sea útil.

La problemática social y la degradación medioambiental provocada por los incendios forestales, demanda una actuación urgente y directa sobre el problema. La investigación en Europa sobre el tema no tiene más de 20 años, y aunque la línea que separa la investigación básica de la aplicada en este campo es muy fina, es una sensación generalizada entre los investigadores de esta rama la falta de coordinación y unicidad de objetivos entre los diferentes grupos de investigación.

Mejorar estas condiciones de trabajo llevaría a un incremento sustancial de productividad, en diferentes sentidos:

- Coordinación, para evitar duplicidad, trabajos repetidos, líneas estériles.
- Integración en sistemas o equipos de investigación.
- Evitar la diseminación de resultados, facilitando su verificación, uso y operatividad.

Otro aspecto clave es optimizar la financiación, tanto en fuentes españolas como europeas. La Comunidad Europea ha financiado dentro de sus programas STEP, EPOCH y ENVIRONMENT, seis proyectos en el período 1991-93, con 4.555.000 ECU, con participantes de 8 países.

De estos seis proyectos, cuatro son sobre comportamiento del fuego, uno sobre efectos posteriores sobre la vegetación y otro sobre prevención.

El programa ENVIRONMENT considera específicamente el tema de los incendios forestales dentro del área “Riesgos Naturales y Tecnológicos”, con los siguientes campos de trabajo:

- Física de los incendios forestales.
- Relaciones entre aspectos del régimen de fuegos.
- Modelización del comportamiento del fuego.
- Impacto y gestión del fuego.
- Prevención y rehabilitación.

Las etapas que deben integrarse y comunicarse, partiendo del objetivo de satisfacer las necesidades reales generadas por la lucha contra los incendios forestales son:

- Previsión de incendios y educación medioambiental.
- Modelización y lucha contra incendios forestales.
- Paliación de efectos posteriores.

Sin esta integración, resulta difícil obtener resultados eficaces y aplicables al problema real y a la lucha diaria.

Desde los comienzos de la integración comunitaria se ha enfatizado la necesidad de una acción coordinada a escala europea en el ámbito de la investigación y el desarrollo tecnológico aplicados a los incendios forestales. Una prueba de ello es el gran número de proyectos que la Comunidad Europea ha financiado [García, 97], además de los que están llevándose a cabo en la actualidad.

Por la relevancia como “estado del arte” para este trabajo, se hará referencia tan solo a aquellos subproyectos seleccionados por el proyecto NOW.

En la Tabla 1 se muestra de modo resumido las características esenciales de los subproyectos europeos que se han desarrollado en el marco del proyecto NOW:

Nombre del Proyecto	Dominio de aplicación	Principales funciones	Tecnologías utilizadas	Resultados explotables
CHARADE	Lucha contra incendios	Evaluación de la situación. Gestión de recursos. Planificador interactivo. Módulo cartográfico	Análisis de tareas. URIAH(GIS). Entorno Spoke. Razonamiento con restricciones. Razonamiento basado en casos.	Prototipo. Análisis de tareas. Metodología.
TRACE	Lucha contra incendios	Evaluación de la situación. Gestión de recursos. Planificador interactivo Módulo cartográfico.	Análisis de tareas. MapObjects(GIS) C++. Razonamiento con restricciones.	Fin del proyecto: diciembre de 1998. 3 prototipos (uno genérico y 2 aplicaciones).
CARICA	Lucha contra incendios	Adquisición de conocimiento. Basado en cartografía. Extracción de conocimiento de informes. Base de datos basada en casos. Explotación del conocimiento para el entrenamiento de los bomberos.	URIAH (GIS). Entorno Spoke. Lisp. Razonamiento basado en casos.	Prototipo.
MEFISTO	Lucha contra incendios	Herramienta para la toma de decisiones en simulaciones en tiempo real de incendios forestales	GIS. Fortran.	Modelo de propagación del fuego. Modelo de la

		incendios forestales. Base de datos geográfica actualizada. Consulta y visualización de anteriores incendios. Ejecutar el simulador de incendios forestales.		energía del viento. Prototipo.
FIREMEN	Lucha contra incendios	Evaluación del riesgo. Soporte inteligente para la preparación de mapas de riesgos de incendios. Soporte para los gestores responsables de la protección del bosque de la planificación y uso óptimo de los recursos.	Visual basic 3.0. Visual C++1.1. Teledetección. Imágenes.	Estrategia basada en el conocimiento para el análisis de datos geográficos diversos (intérprete DSS). Prototipo.
EUCEP	Lucha contra incendios	Estudio de la viabilidad del proyecto CHARADE en España.		Estudio de viabilidad.

Tabla 1: Soluciones propuestas en el proyecto NOW.

De lo expuesto en la Tabla 1 se deduce, en primera instancia, la importancia que la Comisión Europea concede a los proyectos realizados en el campo de la protección civil, lo cual confirma el interés que continúan despertando los sistemas para la lucha contra incendios.

Tras esta consideración genérica, se describen algunos de los aspectos técnicos de cada uno de los proyectos a partir de las fichas resumen de las que se dispone de ellos. Una de las primeras observaciones importantes que se deben señalar, común a todos los proyectos, es la ausencia de una aplicación comercial que figure como resultado de alguno de los proyectos. Este hecho puede resultar llamativo si se tienen en cuenta los recursos destinados a cada una de las iniciativas presentadas. Sin embargo, se trata de algo completamente coherente con la naturaleza realmente compleja del problema que se trata de resolver.

- **CHARADE:** Se trata de un Sistema de Soporte de las Decisiones para la lucha contra incendios forestales, preparado para trabajar en un centro de control de una provincia italiana. El sistema consiste en una serie de módulos

que utilizan razonamiento basado en casos y razonamiento con restricciones, un módulo para planificar la intervención y un módulo para evaluar la situación que incluye un simulador de la propagación del fuego en función del punto geográfico donde se produjo la emergencia.

- **CARICA:** El objetivo de este proyecto consistía en evaluar la idoneidad del módulo de razonamiento basado en casos diseñado en el proyecto CHARADE. Este proyecto generó un sistema para introducir las intervenciones realizadas por los servicios contra incendios del sur de Francia y analizar las decisiones tomadas, siendo de utilidad como sistema de entrenamiento.
- **MEFISTO:** El objetivo del proyecto MEFISTO era diseñar un sistema para el soporte de decisiones para la simulación en tiempo real de incendios forestales. Este sistema sirve como ayuda en la evaluación de riesgos de incendios y para planificar la ubicación de recursos.
- **TRACE:** El principal objetivo del proyecto TRACE era realizar la transferencia de la tecnología CHARADE, desarrollando una plataforma genérica para gestionar emergencias. La principal aportación de este proyecto consiste en haber llevado a cabo una transferencia de tecnología en países de Europa del Este y Central.
- **FIREMEN:** El proyecto FIREMEN ha diseñado un sistema para el soporte de decisiones para riesgos de incendios basado en la evaluación combinada de datos captados por sensores remotos e información geográfica. Además utiliza un modelo de la propagación del fuego junto con una serie de herramientas que facilitan el trabajo de las unidades de intervención.
- En cuanto al proyecto EUCEP, se trata de un documento que realiza un estudio sobre la posible aplicación de la tecnología CHARADE a nuestro país. Este estudio deberá ser actualizado y ampliado.

6. Marco de la tesis.

El origen del interés del autor por esta problemática, se sitúa en los trabajos llevados a cabo durante el periodo comprendido entre 1984 y 1992 como operador informático y de radiocomunicaciones en la Central de Lugo del Servicio de Montes (1984-1989) y Servicio de Defensa Contra Incendios Forestales (1990-1992) de la Consellería de Agricultura de la Xunta de Galicia, durante las campañas de Alto Riesgo.

Desde 1990, el autor investiga en el marco de diferentes proyectos, temas relacionados con la gestión de la vaguedad e incertidumbre en los Sistemas Expertos (Lenguajes Formales Fuzzy para el tratamiento de Conceptos Vagos, financiado por la Xunta de Galicia por el periodo 1.995-96. La gestión de la incertidumbre en los Sistemas Expertos: Aproximación Lingüística e Interfaces Hombre-máquina, de la D.G.I.C.Y.T., 1993-95. Fundamentos de Sistemas Expertos: Formalización y Gestión de la Vaguedad en el Razonamiento Aproximado, financiado por la Xunta de Galicia por el periodo 1.991-92, etcétera).

En 1995, el autor establece una colaboración con el *Institute of Atmospheric Optics* (TOMSK) e *Institute of Computing Techniques* (NOVOSIBIRSK), SB Russian Academy of Sciences, SIBERIA, RUSIA, desplazándose a Siberia para observar *in situ* las técnicas de gestión y lucha contra los incendios forestales en el bosque más extenso de la tierra. Al mismo tiempo, colabora en el planteamiento de un Sistema de Teledetección y proceso Experto de datos sobre incendios forestales con INSA (Ingeniería y Servicios Aeroespaciales, INTA), filial de NASA en España.

Todo ello ha sido realizado en el marco de trabajo del Departamento de Lógica de la Universidad de Santiago de Compostela (Grupo dirigido por el Prof. Alejandro Sobrino), el Departamento de Ciencias de la Computación e Inteligencia Artificial de la Universidad Politécnica de Madrid, donde el autor ha realizado el Master y posgrado en Ingeniería del Conocimiento, así como su programa de Doctorado y el Departamento de Informática de la Universidad de Castilla-La Mancha, donde el autor desarrolla su actividad docente e investigadora.

Capítulo 2

Categorías Prototípicas Deformables Borrosas.

En este capítulo se introduce una noción de prototipo presentada por Zadeh en 1982. Esta teoría fue formulada como reacción a la “teoría ingenua de prototipos”, propuesta desde el ámbito de la Psicología Cognitiva. La teoría de prototipos clásica en psicología se debate en el nivel más general de la categorización conceptual, que tiene como objetivo discutir cómo los humanos conceptualizamos la realidad agrupando objetos en cuanto a sus propiedades o características.

Apoyándose en estos dos pilares, la teoría de prototipos de Zadeh y la clásica de categorización en psicología, se presenta un nuevo concepto: El de “**Categorías Prototípicas Borrosas**”, al que se aludirá genéricamente como Prototipos Borrosos y que es el que servirá de base y objetivo para toda la tarea de Adquisición de Conocimiento automático.

Posteriormente, se muestran las características y dificultades del proceso de Adquisición de Conocimiento, presente en cualquier metodología para el desarrollo de Sistemas Basados en el Conocimiento y se propone una extensión del KDD (*Knowledge Discovery on Databases*): el **DCPB** (Descubrimiento de Conocimiento Prototípico Borroso) como mecanismo para conseguir el objetivo de generar *Prototipos de datos* (subconjuntos o nuevos conjuntos de datos lo suficientemente representativos como para poder resumir o asimilar el comportamiento de todos los restantes datos al de ellos), basados en el concepto de **Categorías Prototípicas Borrosas**, utilizando, entre otras, técnicas de *clustering*, aprendizaje por inducción y *Data Mining*.

Se investiga la noción de prototipo deformable, noción recogida del ámbito del Reconocimiento de Patrones, que, combinada con la noción de Categorías Prototípicas Borrosas proporcionan un mecanismo para superar un problema que frecuentemente se produce en la aplicación de los Sistemas Basados en el Conocimiento: Asimilar el comportamiento de una situación real a un prototipo previamente establecido. Se propone realizar el proceso inverso: Deformar las **Categorías Prototípicas Deformables Borrosas** hasta que describan la situación real.

Para ello, se utiliza el razonamiento no-monótono, haciendo un especial hincapié en el razonamiento con marcos, temporal y borroso no-monótonos, que serán las herramientas para realizar la equiparación (*Pattern Matching*) de un caso real con las Categorías Prototípicas Deformables Borrosas y para deformar éstas para evaluar dicha situación real.

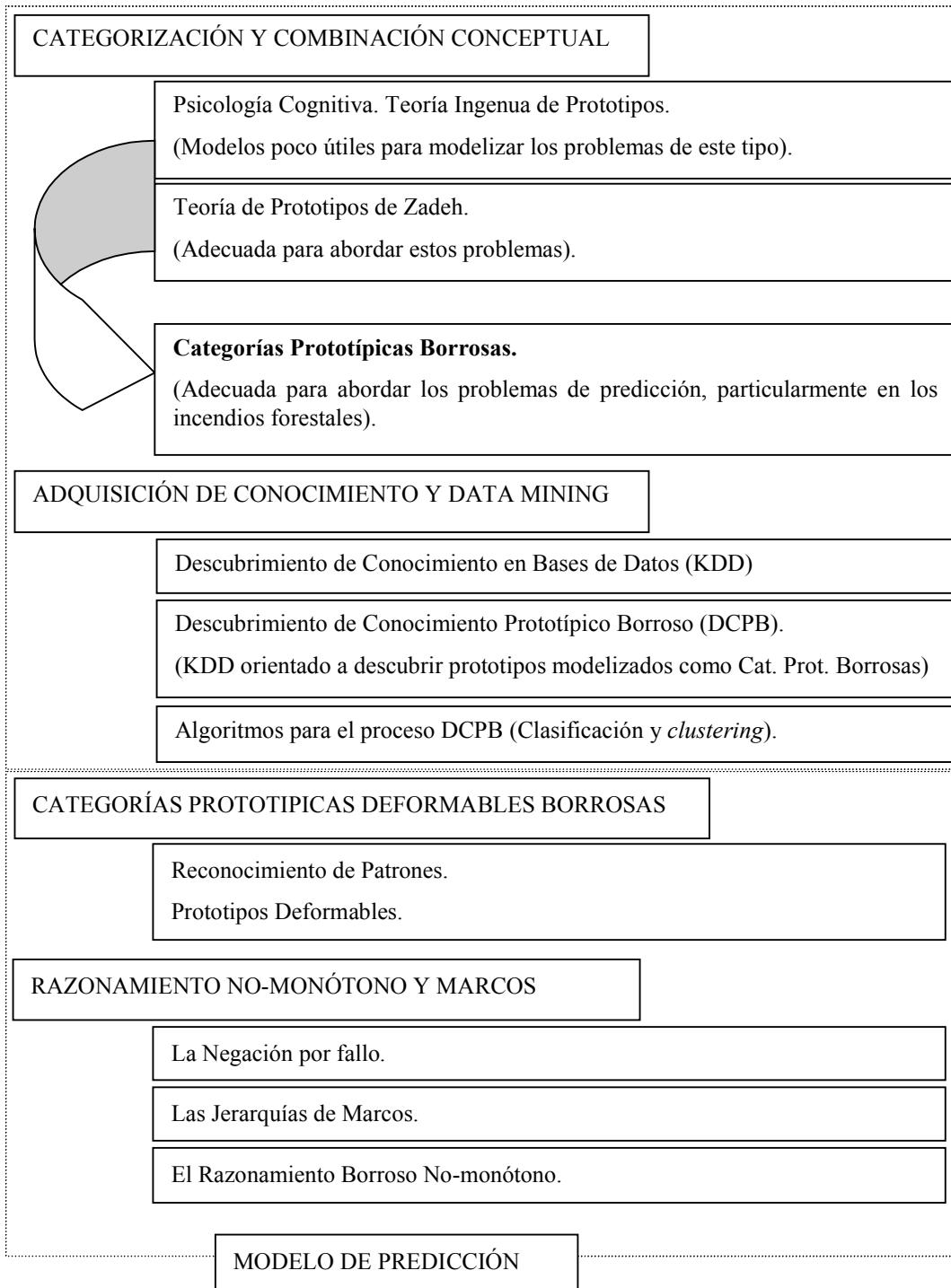


Figura 2: Diagrama de la estructura del capítulo.

1. Categorización conceptual. La noción de prototipo de Zadeh. Categorías Prototípicas Borrosas.

Para que una conducta sea considerada como resolutoria de problemas debe cumplir una condición: debe implicar alguna respuesta nueva para el individuo, no valen recuerdos de respuestas previamente aprendidas, y la solución (en cuanto a método, no a resultado) debe poder aplicarse a la resolución de problemas de similares características. Los psicólogos han utilizado estas premisas principalmente en sus experimentos con animales. Pero en el estudio de los seres humanos se han planteado tareas que van desde anagramas y pasatiempos a razonamientos matemáticos y lógicos, por lo que resulta difícil llegar a conclusiones generales sobre el modo de resolver problemas. Además estos experimentos se ven muy influenciados por condiciones y variables introducidas por el experimentador.

La mayoría de los primeros experimentos en este campo fueron realizados o bien enfatizando las características perceptivas de la situación problema y logrando una intuición (Gestalt), o en el marco de la teoría del aprendizaje Estímulo-Respuesta.

En cuanto a las características de la tarea, parece que es más probable que los problemas sean resueltos si todos los componentes de la solución están visibles, especialmente en los animales ya que los humanos podemos representar lo que falte mediante lenguaje natural, ya sean palabras (en caso de existir) u oraciones. También condiciona la solución la disposición de los objetos necesarios; Duncker [Duncker, 45] realizó una serie de experimentos ya clásicos sobre este tema.

Otra variable importante en la solución de problemas es la complejidad de los mismos. La psicología contempla el grado de abstracción y el grado de definición de la meta deseada como variables esenciales de la complejidad. Cohen [Cohen, 79] compara tres problemas: “Encontrar la raíz cuadrada de 169” (Punto de partida y meta bien definidos), “Reparar un coche” (Meta bien definida, estado inicial mal definido) y “Diseñar un buen centro de ciudad” (Meta mal definida).

Permítase un pequeño inciso: El modo de analizar la complejidad de los problemas en las Ciencias de la Computación contempla propiedades formales

cuantificables que da la impresión que difieren diametralmente de las expresadas por la Psicología. Para los problemas que se abordan en Ingeniería del Conocimiento, parece adecuado analizar la complejidad desde ambos puntos de vista (el de la Psicología y el de las Ciencias de la Computación) debido a la naturaleza imprecisa y/o incierta de muchos de los datos manejados.

El tipo de información que se presenta en el problema inicial y la clase de pruebas que hay que utilizar para resolverlo también pueden jugar un papel importante en la resolución de problemas. Un ejemplo significativo es el uso de la *información negativa*. Los psicólogos han descubierto que habitualmente es más difícil (más lento) evaluar oraciones negativas que positivas. Del mismo modo, a los sujetos a los que se les presenta una *evaluación o descubrimiento de reglas*, tienden a elegir casos de confirmación en vez de sucesos de falsación para comprobar sus hipótesis, aunque muchas veces la información negativa es más esclarecedora.

Para explicar por qué es menos fácil utilizar la información negativa que la positiva, los psicólogos han recurrido a varios tipos de argumentaciones. Se habla de una necesidad de una recodificación *sintáctica* (una *destransformación*) antes de la comprensión, aunque parece que lo más posible es que tenga que haber una recodificación *semántica* (inclinación hacia la información afirmativa y hacia la respuesta afirmativa en vez de hacia la negativa). También se han dado otras explicaciones en términos de connotaciones *emocionales* (proposiciones negativas relacionadas con las prohibiciones y frustraciones lo que provoca ansiedad e inhiben la respuesta). Pero la explicación que más interesa para este trabajo es la que se realiza en términos de *ambigüedad*: en muchas ocasiones la información que comunica una proposición negativa es incompleta o vaga. Por ejemplo, si se dice “el policía no detuvo al coche” se pueden entender cosas como que el policía indicó al coche que siguiera o el policía detuvo a un camión o alguna otra persona detuvo al coche ...

Otro foco de interés de la Psicología ha sido el estudio del razonamiento deductivo e inductivo. En general, los psicólogos dan información acerca de errores del razonamiento (por ejemplo divergencias con los principios deductivos

de la Lógica Formal), aunque también se han hallado pruebas de razonamiento lógico, incluso en sujetos sin un adiestramiento formal.

La causa de estos errores puede estar motivada por las actitudes personales, las *creencias* y las *emociones*, que influyen en la facilidad con que se aceptan las conclusiones lógicas. Otra posible causa puede ser el “efecto atmósfera” [Woodworth, 35] que consiste en la extracción de conclusiones basándose en la impresión global producida por las premisas, pero esta idea ha sido muy cuestionada en toda la Psicología posterior, que defiende que las causas de los errores son producidas porque las inferencias son mayoritariamente inductivas y no deductivas. Cuando se manejan conceptos concretos se suelen obtener mejores resultados; por ejemplo, es más fácil que una persona partiendo de “Todos los A son B” concluya que “Todos los B son A” que si se le dice “todos los leones son mamíferos” concluya que todos los mamíferos son leones.

La *experiencia* juega un papel fundamental en los procesos de razonamiento, ya que permite la transferencia de principios y de destrezas a una situación “nueva”. Pero la experiencia puede ser perjudicial cuando se transfieren estos principios y destrezas a situaciones inapropiadas. Lo mismo ocurre con los métodos que han dado buenos resultados en el pasado, que pueden obstaculizar la solución de problemas nuevos. Un ejemplo clásico tanto para la Psicología como para la Inteligencia Artificial es el de las “jarras de agua” [Luchins, 42], en el que el uso de un método que funciona en muchos casos, en otros lleva a soluciones susceptibles de ser mucho más simples. Se puede usar la experiencia en la resolución de problemas pero no estar dominado por ella.

1.1. *Categorización conceptual.*

La *Formación de Conceptos* es el proceso por el que se llega a responder a los estímulos no en cuanto a acontecimientos individuales sino como miembros de una clase, clasificándolos (separándolos en distintas clases) y categorizándolos (estableciendo jerarquías en las mismas).

La categoría a la que pertenece el estímulo viene determinada por características que varían también en complejidad y grado de abstracción. La formación de

conceptos puede ser “abierta” en vez de “cerrada”, según haya o no un único modo correcto de clasificar los estímulos, por lo tanto el sujeto debe idear uno o más criterios de clasificación. En Ingeniería de Conocimiento es muy frecuente el problema de manejar taxonomías según varios criterios, por ejemplo, en problemas de *clustering*, reconocimiento de patrones, gestión inteligente de bases de datos y en general, cualquier clase de problemas de clasificación.

El tipo de concepto que debe manejarse juega un papel muy relevante, si es que se conoce *a priori*. Los conceptos *conjuntivos* se aprenden mucho más fácilmente que los *disyuntivos* [Bruner, 56], incluso en sujetos muy inteligentes y sofisticados. Para los psicólogos, un *concepto conjuntivo* es aquel en el que todos los miembros de la clase tienen una o más características en común; un *concepto disyuntivo* representa una clase de estímulos cuya pertenencia a la clase viene determinada por el hecho de que posean una de varias características. Dos miembros de una misma categoría disyuntiva pueden ser totalmente distintos y sólo pueden ser identificados por casos negativos del concepto, con las dificultades que esto representa y ya ha sido esbozado anteriormente.

Los experimentos que los psicólogos han utilizado para definir la conducta conceptual han sido, tradicionalmente, muy limitados y artificiales (en el sentido de preparados y manipulados por el experimentador), por lo que resulta difícil extraer conclusiones sobre la conducta conceptual “cotidiana”. Más recientemente se ha intentado hacer una investigación más realista de la formación de conceptos, saliendo del laboratorio. Se ha intentado analizar la “memoria semántica”, el modo en que se clasifica y ordena la información adquirida. Pero el desarrollo que más interés suscita para este trabajo es la investigación sobre la estructura de los conceptos naturales, introducida por Rosch [Rosch 75, 77] y sus colaboradores, quienes proponen que la mejor manera de definir la formación de conceptos es concebirla en términos de la construcción de un *prototipo* que incorpore los rasgos que, en distintos grados, comparten casi todos los casos del concepto: el *prototipo* es, como si dijéramos, un “caso ideal”, en principio no hipotético sino real, con el que pueden compararse los posibles nuevos casos, y con el cual los demás casos tienen un grado diverso de parecido, que los psicólogos han manejado

habitualmente en el sentido de “distancia” lineal, si un gorrión es un prototipo de la clase pájaro, es decir, alcanza 10 en una valoración decimal, se podría decir que un murciélagos alcanza 5, y un pingüino 3, y sus distancias respectivas a un prototipo de su clase son 5 y 7 (nótese que no se alude propiamente a la clase “pájaro” sino a una característica representativa de la misma “animal que vuela”; Si se aludiese a la de “incubación externa”, el murciélagos ya no la tendría).

En los experimentos que ha llevado a cabo este grupo, los sujetos evalúan el grado en que un caso “encaja” en su idea del concepto. También han realizado *experimentos de facilitación*, en los que se investiga el efecto facilitador de presentar un nombre prototípico. Así mismo, se han realizado experimentos de *evaluación del tiempo*, en sentido de estudios evolutivos (procesos que involucran una evolución temporal y no tratan el tiempo como una variable estática más).

Los enfoques de este tipo están ampliando y cambiando el estudio de la formación de conceptos, y parece probable que habrán de proporcionar datos e hipótesis que tendrán relevancia para el estudio, no sólo del pensamiento, sino también de la memoria y de la percepción.

En ocasiones, un conocimiento profundo de estos estudios del ámbito de la Psicología puede ser útil a la Ingeniería del Conocimiento. La aplicación de los resultados y los modelos obtenidos podrá permitir implementar herramientas más cercanas a los modelos humanos de inteligencia. En este trabajo se pretende conjuntar conceptos (principalmente el de *prototípico*) del ámbito de la Psicología con herramientas formales que proporciona la lógica (especialmente la borrosa y la temporal) para desarrollar un modelo computable de razonamiento, que sea útil en procesos de predicción.

Para concluir esta visión de la categorización conceptual es conveniente comentar las seis más relevantes cuestiones teóricas y prácticas que centran la investigación y el debate en esta materia:

1.1.1. Arbitrariedad de categorías. ¿Hay razones a priori para dividir los objetos en categorías, o esta división es inicialmente arbitraria?.

La naturaleza no arbitraria de las categorías.

La idea de que dividir los objetos del mundo real es en principio arbitraria, podría tener sentido sólo si los atributos del mundo constituyesen un conjunto total, es decir, si cualquier combinación de valores de los atributos tuviese la misma probabilidad de ocurrir.

Supóngase, por ejemplo, que se quieren clasificar animales considerando los siguientes atributos con sus posibles valores:

Pelo: Piel, plumas.

Apertura oral: Boca, pico.

Forma principal de locomoción: Volando, caminando.

Si se considerase esta idea de conjunto total, se podrían diferenciar ocho tipos de animales:

Piel, boca, camina.

Piel, boca, vuela.

Piel, pico, camina.

Piel, pico, vuela.

Plumas, boca, camina.

Plumas, boca, vuela.

Plumas, pico, camina.

Plumas, pico, vuela.

No es obvio cómo asignar estas hipotéticas criaturas a categorías. Realmente sólo la primera (mamíferos) y la última (pájaros) contienen a la mayoría de las especies que existen en realidad en la tierra.

Se podría comenzar a trabajar con el concepto de *taxonomía*. Los objetos pueden ser categorizados en varios niveles jerárquicos diferentes. Antropólogos y

psicólogos han llegado a la conclusión de que los niveles fundamentales y más eficientes a la hora de categorizar son los más básicos. Rosch y sus colaboradores [Rosch, 76] han mostrado que el nivel básico es el más general, debido fundamentalmente a tres factores:

- a. Las personas usamos acciones motoras similares para interactuar con los miembros de dichas categorías.
- b. Los miembros de la categoría tienen una forma genérica similar.
- c. Una imagen mental puede reflejar toda la categoría.

Se puede considerar esta categorización a nivel básico como un proceso básico. Muchos trabajos del ámbito de la Psicología muestran que sin algún tipo de categorización, un organismo no podría interactuar con la gran cantidad de objetos y eventos que experimenta. Los humanos, desde los primeros meses de vida, hacemos uso de ese proceso, por ello es considerado básico.

1.1.2. Equivalencia de los miembros de la categoría. ¿Son todos los miembros igualmente representativos de la categoría, como muy frecuentemente es asumido?

La no equivalencia de los miembros de una categoría.

Si las categorías son vistas como establecidas por criterios de pertenencia de necesidad y suficiencia, entonces cualquier miembro de una categoría podría ser cognitivamente equivalente a cualquier otro.

Pero tras muchos experimentos de antropólogos y psicólogos se ha demostrado que existen grados de representatividad para muchas categorías, sobre todo semánticas. La representatividad se define operacionalmente mediante el *ranking* subjetivo de cuan bueno es un ejemplo de esa categoría [Rosch, 75]. Las principales medidas usadas en investigación psicológica para medir la representatividad de un elemento con respecto a una categoría son las siguientes:

- a. *Velocidad de proceso.* A un sujeto se le pide que responda lo más rápidamente posible a una pregunta de la forma ¿Un (elemento) es un (nombre de la

categoría)?. Cuando los elementos son los considerados prototípicos las respuestas son más rápidas.

- b. *Orden y probabilidad en la producción de ejemplos.* Normalmente existe una correlación en la frecuencia al mencionar determinados ejemplares de una categoría, que suelen ser los más representativos, sobre todo para categorías semánticas.
- c. *Uso de términos de la categoría en el lenguaje natural.* Determinadas “cercas semánticas” suelen dar pistas sobre la representatividad de una determinada clase. Por ejemplo se puede decir que “un pingüino es *técnicamente* un pájaro” o “un gorrión es *realmente* un pájaro”.
- d. *Asimetrías en las relaciones de similaridad entre ejemplares de una categoría.* Habitualmente los elementos menos representativos se suelen considerar similares a los más representativos y no al contrario.
- e. *Aprendizaje y desarrollo de las categorías.* Dos factores relevantes en cuanto a los grados de representatividad: El primero es que la pertenencia a una categoría es establecida por las personas comenzando por los ejemplares más representativos y terminando por los menos. El segundo es que las categorías son aprendidas más fácilmente y más correctamente si la exposición inicial se hace sólo con ejemplares representativos.

Pero, ¿Qué es lo que hace a algunos objetos más representativos de su categoría que otros?. Lo normal es considerar la correlación de los atributos. En el ejemplo de las criaturas anterior, hay bastantes pájaros que “medio vuelan”, algunos mamíferos que vuelan (murciélagos) y más criaturas intermedias. Por ello, el grado de representatividad tiene que ver con el nivel en la taxonomía. Los elementos más representativos están asociados con el nivel básico que ya se ha comentado.

Rosch y Mervis mostraron que dado un conjunto de atributos del mundo real, los elementos que tienen mas atributos en común con los otros elementos de su categoría también son los que tienen menos en común con otras categorías.

El fenómeno de la no equivalencia de la representatividad de los elementos de una categoría también ha sido estudiado desde el ámbito de la sociología. Se ha tratado el concepto de *estereotipo* e incluso se han implementado mecanismos (por ejemplo basados en inferencia Bayesiana) para determinar que atributos deberían ser incluidos en un estereotipo.

1.1.3. Determinación de la pertenencia y representación de las categorías. ¿Está la pertenencia a las categorías especificada mediante condiciones necesarias y suficientes?. ¿Están bien definidos los límites de las categorías?.

La indeterminación de la pertenencia y la representación de las categorías.

La variación de los miembros de una categoría en cuanto a representatividad es esencialmente un conjunto de descubrimientos empíricos. ¿Qué implicaciones tiene esta afirmación con respecto a la naturaleza de las categorías?. Dos consideraciones:

- a. Una cuestión relativamente empírica sobre si los límites de las categorías están bien definidos o son borrosos. Se han utilizado varios experimentos para demostrar que los límites de las categorías no están bien definidos. Unos se basan en desacuerdos entre sujetos sobre la pertenencia a las categorías de ejemplares con poca representatividad. Otros en que los elementos pobres de algunas categorías poseen atributos de otras.
- b. Una cuestión teórica sobre cómo de bien definida debe ser considerada la categoría “representación” en sí misma. Si uno cree que las categorías consisten en determinar criterios de necesidad y suficiencia, entonces puede desarrollar un modelo que intente explicar la representatividad y el carácter indeterminado de los límites mediante procesos que operen en una determinada representación.

En lógica borrosa se usa otra estrategia, donde los conceptos son representados mediante una especificación determinada de los atributos definidos, pero los atributos (se presupone que no son conceptos), son tratados como borrosos.

Estos modelos no cambian con la variación en los descubrimientos empíricos [Smith, 81].

1.1.4. La naturaleza de la abstracción. ¿Cuánta abstracción se requiere, sólo memoria para los elementos o abstracciones de alto nivel, de conocimiento general?.

En el paradigma clásico de formación de conceptos, un concepto es una abstracción que consiste en un conjunto de características definidas y las relaciones entre ellas. Todos los modelos de categorización incluyen abstracción (es decir, la forma en que los sistemas cognitivos actúan de forma creativa cuando aprenden las categorías y usan la información resultante para clasificar nuevos casos). Esta creatividad es de dos tipos:

- a. Determinación de cuáles son los elementos esenciales y cuáles los irrelevantes en una determinada situación.
- b. Creación de una información de orden superior que no es dada por ningún ejemplar en particular

En el paradigma clásico de formación de conceptos, la abstracción de elementos esenciales pasa por aprender qué atributos son relevantes y la creación de información de orden superior se basa en aprender las relaciones lógicas entre esos atributos.

Cualquier modelo que incluya representación de características posee creatividad del primer tipo. El tipo de nueva información de orden superior generada varía para los diferentes modelos. La menor complejidad computacional la requieren los modelos de prototipos basados en la tendencia central, por ejemplo las medias, modas o valores ideales basados en características perceptuales. Algunos modelos incluyen información de resumen e información de los ejemplares. Para esta información de resumen, estos modelos necesitan abstracción de elementos “esenciales” para generar características y la creación de nueva información en forma de pesos de las características (para computar de forma que las características sirvan para determinar la pertenencia a la categoría de nuevos casos).

Estos modelos de abstracción con atributos simples, reglas o prototipos son criticados por los que creen que las abstracciones de orden superior y conocimiento general son más extensas que las categorías individuales usadas en los procesos de categorización. Suelen tratar las categorías como parte de esquemas muy generales.

En el ámbito de la computación se han tratado las categorías y los procesos de categorización como inseparables del conocimiento del mundo (o dominio) y los procesos de inferencia usados sobre ese conocimiento [Schank, 77], [Winograd, 72].

1.1.5. Descomposición de categorías en elementos. ¿Es razonable explicar los objetos mediante su descomposición en cualidades elementales?.

La capacidad de descomponer las categorías en elementos.

La mayoría de los trabajos sobre representación y procesado de categorías asumen que pueden ser descompuestas en cualidades más elementales. Por ejemplo, en aplicaciones de *pattern recognition* los modelos que se presentan o se basan en análisis de características o en un emparejamiento holístico (integral) con un modelo.

Dentro de los desarrollos empíricos que se han llevado a cabo para examinar la naturaleza y papel de la descomposición, se pueden destacar tres grupos:

- La categorización ha sido investigada por tipos de estímulos que no poseen elementos obvios a nivel cognitivo (por ejemplo el color, que no es fácilmente descomponible).
- Un segundo desarrollo que ofrece la oportunidad de ver las categorías como conjuntos (como un “todo”) es la posibilidad de representación espacial de las estructuras dentro de y entre las categorías, con técnicas por ejemplo de espacios multidimensionales.
- Como tercera aproximación se ha usado el concepto de prototipo y los grados de representatividad, que sugieren un procesamiento holístico.

La gran mayoría de argumentos acerca de la descomposición hacen referencia específica al nivel de abstracción al cual una clase particular de descomposición puede o no ocurrir. Algunos elementos son incluidos como “primitivos”, normalmente por defecto. Como algunos elementos no son descomponibles, muchos mecanismos de categorización incluyen un componente holístico explícito. Otra posibilidad es que un nivel de abstracción dado podría ser un nivel básico y potencialmente holísticamente percibido, aun cuando otros niveles requieran más mecanismos analíticos.

Es importante destacar la controversia que se da en Lingüística e Inteligencia Artificial sobre la descomposición léxica, donde hay problemas para discernir qué etiquetas lingüísticas deben ser consideradas como “todos” y cuáles pueden ser elementos.

A modo de conclusión, se han tenido dos puntos en cuenta en la discusión sobre descomponibilidad:

- Aunque la tendencia de los modelos cognitivos es descomponer automáticamente, la evidencia de procesos holísticos de algunos estímulos o en algunos estados sugiere que se debe meditar más acerca de estos modelos.
- Los hallazgos sobre descomposición parecen ser dependientes del nivel que se considera. Esto implica que se necesitan algunos principios generales de descomposición.

1.1.6. La naturaleza de los atributos. ¿Cuáles son las características de esos atributos en los que se han de descomponer las categorías?.

Para los empiristas ingleses, los atributos se correspondían con sensaciones elementales. La psicología cognitiva moderna considera que cualquier cosa puede ser considerada atributo en un momento u otro. Esto genera anomalías en el uso de partes, relaciones y funciones como atributos. Que una cosa sea categoría o atributo depende del nivel en el que se esté trabajando.

La primera controversia surge en el uso de características *versus* el uso de dimensiones en la representación de las categorías:

- Las *características* normalmente designan propiedades cualitativas y no necesitan ser aplicables a todos los objetos del mismo dominio. Un número grande de características pueden ser incluidas en una representación simple. Hay muchos tipos de representaciones como las listas de características o las descripciones estructurales.
- Por el contrario, las dimensiones se utilizan habitualmente para describir propiedades cuantitativas. A cada objeto se le asigna algún valor de cada dimensión usada para describir el dominio. Una representación dimensional ideal incluiría sólo un pequeño número de dimensiones. Hay dos tipos de representación dimensional: métrica y no-métrica. A menudo usan técnicas de análisis multidimensional.

Características, transformaciones y dimensiones no-métricas pueden ser integradas de una forma razonable en la misma representación.

Otra aproximación se refiere a cómo unos tipos de atributos combinan con otros. Se pueden distinguir dos tipos de combinaciones de atributos:

- *Separables*: Si se perciben en términos de sus atributos separados, por ejemplo forma y tamaño.
- *Integrales*: Si los atributos no pueden considerarse de forma separada, por ejemplo brillo y saturación en una imagen.

Los adultos habitualmente consideramos ambos tipos de combinaciones, mientras que los niños pequeños tienden a tratarlos todos como integrales.

En la filosofía empirista, los atributos eran usados para conectar conceptos al mundo real. Es decir, para conectar el significado (intensivo) de un concepto con el objeto (extensión) al que alude este significado. En psicología de categorización también son usados a menudo los atributos para este propósito. Estos atributos suelen ser generalmente de cuatro tipos:

- Partes.
- Características físicas.

- Conceptos relacionales.
- Conceptos funcionales.

1.1.7. Resumen.

Como resumen de lo hasta aquí expuesto se pueden extraer tres conclusiones fundamentales para el trabajo que aquí se desarrolla:

1. Las categorías están internamente estructuradas mediante grados de representatividad.
2. Los límites de las categorías no son necesariamente nítidos (definidos).
3. Hay una relación muy estrecha entre los *clusters* de atributos y la estructura y formación de las categorías.

1.2. Teoría “Ingenua” de Prototipos.

Como se ha visto, el concepto “*Prototipo*” se refiere a la representatividad de un objeto con respecto a una clase. En este sentido, hay objetos que pertenecen a una misma clase pero no son igualmente representativos de la misma, es decir, aún habiendo elementos que pertenecen a una clase (Terminología de Teoría de Conjuntos) o son miembros de una categoría conceptual (Terminología más cercana a la Psicología Cognitiva), no tienen la misma representatividad para ella.

Rosch no duda de que el murciélagos pertenece a la clase de los pájaros, pero no es tan representativo de la clase pájaro como otros [Rosch, 75], por ejemplo el gorrión; es decir, a ambos animales no se les atribuye normalmente la misma pertenencia a la clase pájaro. Si es presentada una colección de pájaros con diferencias entre sí (vuela/no vuela, granívoro/no granívoro, ...) a un conjunto de sujetos para que los clasifiquen según su representatividad, los catalogarían a todos como pájaros, pero con distinto grado de pertenencia (en el sentido de *distancia lineal* ya explicado) al conjunto de los pájaros.

Aunque muy operativa en lo que se ha denominado “teoría de prototipos ingenua”, que responde básicamente a los estudios de Rosch anteriormente reseñados, esa

concepción de lo que es un prototipo no parece plenamente satisfactoria, no es tan intuitiva como aparenta:

Por un lado, es sabido que “pertenencia” no es lo mismo que “representatividad”: es dudoso que un *autogiro* sea tan representativo del concepto “aeronave” como un *Boeing 747*. Sin embargo, ambos serían admitidos como pertenecientes a la clase de las aeronaves.

Por otro, se intuye que un prototipo, si responde a la forma de adquirir conocimiento y clasificarlo de los humanos, no tiene porqué coincidir con el elemento máximamente representativo de la clase, en caso de que éste exista. En efecto, parece dudoso que si un terráqueo cualquiera le diese una descripción de sus congéneres a un extraterrestre, éste identificase mejor o más rápido como “humano” a un/a “modelo de pasarela” (prototipo) humano, que a un individuo con características normales.

1.3. Lógica Borrosa y Prototipos: Teoría de Prototipos de Zadeh.

Lo hasta el momento expuesto incita a estudiar el concepto de prototipo de Zadeh, que inmediatamente se revela como una noción que puede ser fructífera en su aplicación a diversos temas.

Se pueden distinguir dos definiciones de prototipo para conceptos que se consideran vagos:

1. Desde la psicología, en el marco de una “teoría de prototipos”, un prototipo tiene una doble caracterización:

- Aquel elemento que es máximamente representativo para el concepto;
- El elemento con máxima pertenencia al conjunto que representa al concepto.

Ejemplo clásico ya citado es el de Rosch, que sostiene que un elemento, el “petirrojo”, es un prototipo para un concepto (pájaro) o para una clase (clase de los pájaros). Esta es la acepción más usual de lo que es un prototipo, pero no la única.

2. Zadeh [Zadeh, 1982] resume su desacuerdo con una teoría de prototipos desde la psicología, como respuesta a las propuestas de Osheron y Smith [Osheron, 81][Smith, 84] destacando tres ideas:

- Un objeto puede estar lejos de representar bien a un concepto y sin embargo pertenecer completamente a él [Armstrong, 83].
- Un prototipo casi nunca es un objeto sencillo. Es más bien un esquema borroso que muestra en qué medida distintos objetos representan a un concepto.
- Como consecuencia del punto anterior, la prototipalidad es entonces un problema de grado. Como se ha indicado, para Zadeh es dudoso que el prototipo de un concepto sea un elemento.

Esta concepción es acorde con lo que él considera representación lógica de una proposición que incluye predicados vagos e individuos susceptibles de ser atribuidos a esos predicados.

Si se considera P un predicado vago y s_1, \dots, s_n son distintos elementos susceptibles de caer bajo la extensión de P , la proposición vaga $P(s_1, \dots, s_n)$ puede ser representada por una distribución de posibilidad que por definición es numéricamente igual al grado de pertenencia de los s_1, \dots, s_n a P .

Según el grado de pertenencia alcanzado, se puede clasificar a esos elementos en tres clases: los que son ejemplares buenos, fronterizos y malos del concepto. De esta manera se tiene al conjunto P estructurado en tres partes, de forma que su valor sería la suma de los valores parciales referidos. Pero este sería el valor de P , no del prototipo de P , ($PT(P)$).

$$P = \text{Alto}/_A \text{Buenos} + \text{Medio}/_A \text{Fronterizos} + \text{Bajo}/_A \text{Malos}$$

Para obtener $PT(P)$, se debe seguir el siguiente proceso: Los ejemplares buenos, ya con un grado atribuido, se vuelven a valorar respecto a en qué medida son prototipos de los buenos ejemplares de P . A continuación, se indica en qué grado ese grado de prototipo es alto. Esta operación se repite con los ejemplares fronterizos y malos. La suma (no entendida en un sentido aritmético, sino más

bien como reunión) de todos ellos constituye el valor del $PT(P)$. Por tanto, este no tiene porque ser un valor sencillo, lo cual es corriente en lógica borrosa, ni tampoco tiene que ser identificable con un objeto sencillo.

$$PT(P) = \text{Alto}_{PT(P \text{ Buenos})} + \text{Medio}_{PT(P \text{ Fronterizos})} + \text{Bajo}_{PT(P \text{ Malos})}.$$

La aproximación de Zadeh a lo que debe entenderse por “prototipo” es menos intuitiva que la de una teoría de prototipos desde la psicología, pero más racional; menos acorde con lo que parece ser el uso habitual de este término, pero más próxima a lo que, en un examen detenido, muestra el significado de un concepto prototípico.

El uso de esta noción de prototipo todavía no ha sido suficientemente explotado, incluso dentro de la propia área de creación de teoría borrosa. Así, por ejemplo, la interesante noción de sinonimia, un tipo de relaciones de similaridad de Zadeh usa la noción ingenua de prototipo, no la borrosa. En este trabajo se muestra el uso de una extensión de esta concepción en un sistema real y complejo.

1.4. Categorías Prototípicas Borrosas.

A partir de los estudios de la Psicología Cognitiva y del concepto de prototipo de Zadeh, se está en condiciones de hacer una nueva definición del concepto de prototipo, tomando muchas de las conclusiones a las que ha llegado la Psicología pero dotándolo de un mecanismo formal con herramientas de la lógica borrosa que lo hacen aplicable en Ingeniería del Conocimiento.

El concepto que se presenta es el de **Categorías Prototípicas Borrosas**, que hereda de la Psicología las siguientes características, ya reseñadas con anterioridad:

1. Las categorías están internamente estructuradas mediante grados de representatividad.
2. Los límites de las categorías no son necesariamente nítidos (definidos).
3. Hay una relación muy estrecha entre los *clusters* de atributos y la estructura y formación de las categorías.

Formalmente se plasma de la siguiente forma (ver fig. 3):

1. Debido a que hay una relación muy estrecha entre los *clusters* de atributos y la estructura y formación de las categorías, éstas han de ser descubiertas y organizadas mediante algoritmos de *clustering jerárquico* (en esta instancia del modelo se usará el método de emparrillados, *Repertory Grids*).

Los resultados de Categorías Prototípicas Borrosas y de sus atributos se plasmarán en tablas (con estructura de marcos) que en este trabajo también serán denominadas Prototipos Conceptuales o Prototipos Borrosos.

2. A diferencia con el modelo de Zadeh, las categorías pueden ser tantas como requiera el problema. En el Apéndice 3 de este trabajo se puede ver que esto ocurre en otras aplicaciones del método: INFEDEC 2.1 maneja más de 200 categorías (enfermedades) y el reconocedor de olores mediante sensores de estado sólido maneja 4 categorías de combustión. En el modelo de Zadeh se hacen tres particiones borrosas sobre un determinado concepto (los buenos, limítrofes y malos representantes del concepto).

En el modelo que se propone, las Categorías Prototípicas Borrosas pueden ser tantas como requiera el problema. Es decir, la partición borrosa se realiza sobre el dominio y no sobre cada concepto (cuando el problema lo requiera se puede utilizar la definición de Zadeh para cada concepto).

3. Por último, estos Prototipos Conceptuales se representan mediante un *term-set* donde cada etiqueta se corresponde con una categoría y se representa con un número borroso. El dominio de este *term-set* se corresponde con el dominio del problema, pero representado unidimensionalmente, es decir, agregando las diferentes dimensiones que pueda tener el dominio real (en la instancia del modelo se usarán varios tipos de *medias* de los parámetros normalizados como funciones de agregación).

Para cada problema determinado se define uno o varios contextos que guiarán el proceso de *clustering* mediante el que se obtendrán las Categorías Prototípicas Borrosas.

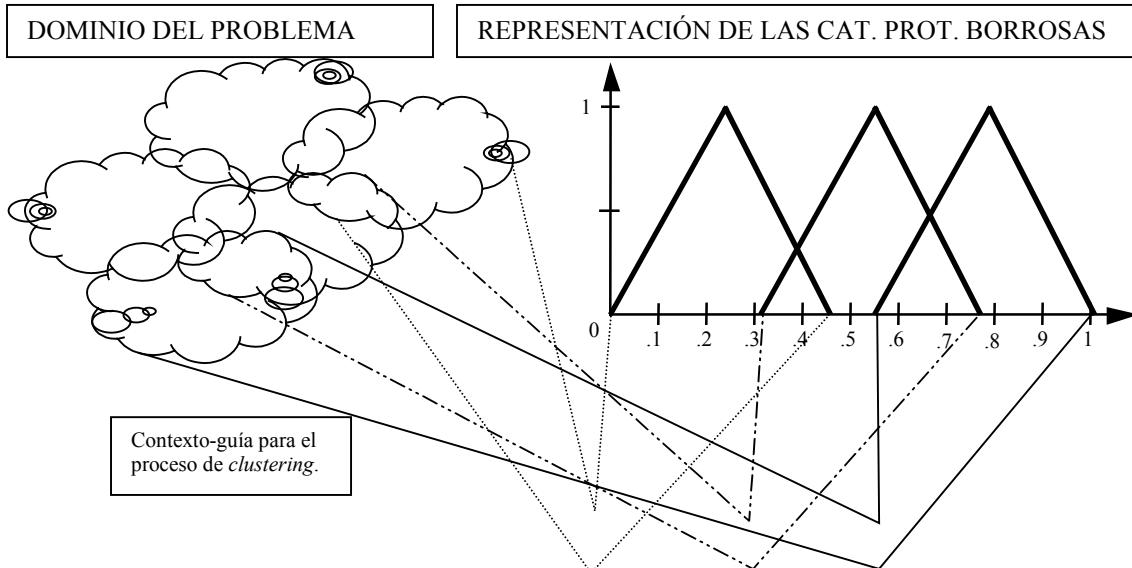


Figura 3: Categorías Prototípicas Borrosas.

1.5. La Lógica Borrosa en el KDD. El proceso DCPB.

Como habitualmente se define en la literatura, el KDD consiste en identificar patrones interesantes y describirlos de una forma concisa y comprensible. Como se aplica con una gran versatilidad, hay un componente muy importante de interacción humana, inherente a los procesos de representación, manipulación y procesado del conocimiento. Esto provoca un creciente auge de la aplicación de tecnologías borrosas en este campo [Pedrycz, 98], principalmente porque los conjuntos borrosos son adecuados para manipular conceptos como granularidad o imprecisión en la información. Como el KDD es una técnica o disciplina que todavía está naciendo, lo que se pretende en este trabajo es establecer un modelo que vincule el KDD con algunas técnicas usuales y otras novedosas del campo de la Lógica Borrosa.

Es previsible un desarrollo acusado del KDD en un futuro próximo, porque indiscutiblemente, la sociedad actual es rica en datos pero pobre en conocimiento (por ejemplo los sistemas de observación de la NASA generan 50 Gb. de datos de imágenes por hora).

De los elementos principales que debe poseer el KDD, la mayoría se pueden cuantificar mediante conjuntos borrosos o utilizar éstos para conseguirlos mejor:

- **Validez:** Relevancia del conocimiento descubierto.
- **Novedad:** Debe describir cuánto se desvían del conocimiento inicial los patrones descubiertos.
- **Utilidad:** Cuán útiles son los patrones descubiertos.
- **Simplicidad:** Cómo se presentan los patrones descubiertos, sobre todo a nivel de complejidad sintáctica y de interpretación.
- **Generalidad:** Cuántos casos de los iniciales son abarcados por los patrones descubiertos.

El proceso de Data Mining, puede ser visto como el descubrimiento de la estructura y eventualmente de la cuantificación de las dependencias funcionales en una tabla relacional X, compuesta por objetos (filas), representados por vectores de números reales.

En este caso, lo que se hace normalmente es usar métodos estándar de *clustering* o *fuzzy clustering* [Bezdek, 91], de los que merece la pena destacar los siguientes aspectos:

- En un principio se han estudiado los niveles de abstracción necesarios para los procesos de clustering desde el punto de vista de la Lógica Borrosa [Bellman, 66]. Enrique Ruspini [Ruspini, 69] introdujo una nueva formalización y perspectiva en el análisis de *clusters*, lo que llevó a nuevas técnicas y algoritmos, como el uso de Conjuntos Borrosos Unimodales [Gitman, 70], Matrices de covarianza borrosas [Gustafson, 79], algoritmos de *fuzzy clustering* geométrico [Windham, 83], algoritmos basados en relaciones borrosas [Tamura, 71] y a partir de ellas un análisis de grafos [Dunn, 74], árboles de decisión borrosos [Chang, 77] y sobre todo los algoritmos ISODATA [Dunn, 73] y *Fuzzy c-Means*, de los cuales ha habido gran cantidad de extensiones hasta nuestros días y que se analizarán a continuación con más detalle.
- Se ha diseñado un teorema de convergencia para los algoritmos ISODATA y *Fuzzy c-Means* [Bezdek, 80 y 87], y también se ha estudiado su consistencia

[Sabin, 87]. Se ha estudiado la validez de estos algoritmos y se han diseñado medidas borrosas para cuantificarla ([Windham, 82], [Xie, 91]).

- Ha habido diversas aplicaciones a problemas muy acotados, como por ejemplo al reconocimiento de caracteres numéricos [Siy, 74] o al reconocimiento de la madurez de los esqueletos a partir de imágenes de rayos-X [Pathak, 86]. Se han comenzado a utilizar redes neuronales borrosas para la solución de problemas de clasificación y *clustering*.
- Las tendencias actuales se encaminan hacia los métodos de optimización, tanto en el preprocesso de los datos [Genther, 97], como en los propios algoritmos de clustering ([Fu, 98], [Jiménez, 97]), con el objetivo de que puedan ser aplicados a los conjuntos de datos de grandes dimensiones que actualmente se manejan. También se optimizan técnicas de redes neuronales o algoritmos genéticos.

La particularización para cada problema del mecanismo de *Data Mining*, lleva a una búsqueda dirigida de regularidades y dependencias en un contexto que el usuario especifica. Esta naturaleza dirigida del proceso se lleva a cabo normalmente mediante término/s lingüísticos específicos, antes de atacar el problema con algoritmos de cálculo muy intenso. Básicamente se pueden considerar dos casos:

Un único atributo representa la variable de contexto, por lo que el término lingüístico que lo representa se representará en un conjunto borroso A:

$$A: Y \rightarrow [0, 1]$$

Siendo Y el universo de discurso de este atributo. Ahora el problema reside en descubrir la estructura de X en el contexto de *A*, donde el contexto en KDD se establece como:

$$\mathcal{A} = \{A: X \rightarrow [0, 1]\}$$

Por ejemplo, si la variable seleccionada es “evolución de la siniestralidad”, el problema del *Data Mining* será:

Descubrir la estructura de X en el contexto {evolución de la siniestralidad = escasa}

Varios atributos forman el contexto. Por ejemplo. A y B dos conjuntos borrosos definidos en Y y Z respectivamente. A estará formado por el producto cartesiano de A x B:

$$A = \{A \times B : Y \times Z \rightarrow [0, 1]\}$$

Descubrir la estructura de X en el contexto {evolución de la siniestralidad = escasa y nº de incendios = bajo}

Además de estas dos formas básicas de contextos lingüísticos hay otras extensiones muy interesantes:

Contexto lógicamente compuesto: {evolución de la siniestralidad = escasa y nº de incendios = bajo o superficie quemada = baja}.

Contexto relationalmente compuesto: {superficie arbolada quemada y superficie rasa quemada son similares}.

Contexto de regresión compuesto: {error del modelo lineal $x_i = f(x_j, a)$ es pequeño}.

Cualquier problema de la forma “Descubrir la estructura de X ...” es de tipo *context-free* y se puede abordar con los métodos clásicos de *clustering*.

1.5.1. Clustering Borroso dependiente del contexto.

Para el KDD resulta fundamental disponer de una guía que ayuda a navegar por los datos y que dirija cualquier mecanismo de búsqueda. Esto previene de encontrar patrones triviales o sin significado. Por lo que se abordarán el *clustering* borroso sensible al contexto y el *clustering* en el espacio lingüístico inducido. Previamente se describe de forma genérica el algoritmo *Fuzzy C-Means*:

Asúmase que x_1, x_2, \dots, x_N son patrones n-dimensionales definidos en R^n . La función objetivo se define como la suma de los errores cuadráticos:

$$Q = \sum_{i=1}^c \sum_{k=1}^N u_{ik}^2 \|x_k - v_i\|^2$$

Con $U = [u_{ik}]$ es una matriz partición, $U \in U$, c es el número de *clusters*. $\|\cdot\|$ es una función de distancia entre x_k y el prototipo (v_i) del i -ésimo *cluster*. U denota una familia de $c \times N$ matrices:

$$U = \left\{ u_{ik} \in [0,1] \mid \sum_{i=1}^c u_{ik} = 1, \forall k \text{ y } 0 < \sum_{k=1}^N u_{ik} < N, \forall i \right\}$$

Formalmente, el problema de optimización se puede expresar de la forma:

$$\min_{U, v_1, v_2, \dots, v_c} Q,$$

con $U \in U$.

El esquema iterativo para evitar mínimos locales o “sillas de montar” en Q es conocido, y se trata con actualizaciones de la matriz de particiones:

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{\|x_k - v_i\|}{\|x_k - v_j\|} \right)^2}, \quad i = 1, 2, \dots, c; \quad k = 1, 2, \dots, N$$

Los prototipos de los *clusters* se obtienen como medias ponderadas de los patrones:

$$v_i = \sum_{k=1}^N u_{ik}^2 x_k / \sum_{k=1}^N u_{ik}^2$$

1.5.1.1. Clustering Borroso sensible al contexto.

La sensibilidad al contexto se le introduce al algoritmo considerando que la variable que condiciona (contexto) toma los valores f_1, f_2, \dots, f_N en los correspondientes patrones. Dicho de una forma más precisa, f_k describe el nivel de integración de x_k con el contexto asumido, $f_k = A(k)$. A actúa como un filtro de *Data Mining* que dirige la atención hacia algún subconjunto de datos.

Si se modifican los requerimientos, se define la familia:

$$U(f) = \left\{ u_{ik} \in [0,1] / \sum_{i=1}^c u_{ik} = f_k, \forall k \text{ y } 0 < \sum_{k=1}^N u_{ik} < N, \forall i \right\}$$

El problema de optimización se puede reformular como:

$$\min_{U, v_1, v_2 \dots v_c} Q,$$

con $U \in U(f)$.

Y el esquema iterativo para la minimización de la función objetivo es:

$$u_{ik} = \frac{f_k}{\sum_{j=1}^c \left(\frac{\|x_k - v_i\|}{\|x_k - v_j\|} \right)^2}, \quad i=1,2,\dots,c; \quad k=1,2,\dots,N$$

Los prototipos de los clusters se obtienen de la misma forma que en el algoritmo original. Es importante señalar que el contexto A tiene un efecto dominante sobre el mecanismo de *clustering*. Si $f < f'$, la población de patrones involucrados en agrupar bajo el contexto f' será menor.

1.5.2. Clustering en un espacio lingüístico inducido.

Casi todos los algoritmos de *fuzzy clustering* operan en el espacio original o normalizado linealmente donde los puntos están definidos. La estructura de la búsqueda puede estar dirigida definiendo marcas lingüísticas (como conjuntos o relaciones borrosas), que convierten los datos originales en un nuevo espacio lingüístico. La granularidad lingüística lleva a cabo una transformación heterogénea del espacio original aplicando mecanismos de expansión o contracción locales. Las marcas se especifican mediante el análisis de datos en términos de su distribución y su granularidad (se convertirá el proceso de *clustering* en uno de clasificación; Por ejemplo, en el caso de la predicción de incendios, una métrica realizará la transformación de los datos obtenidos a través de un procedimiento de *clustering* jerárquico al espacio lingüístico inducido de los prototipos de evolución de la siniestralidad)

1.5.3. Modelado borroso y Data Mining.

El modelado de sistemas (y en particular de sistemas borrosos) y el Data Mining poseen algunas relaciones interesantes [Pedrycz, 98]:

- **Relaciones entre variables bien definidas y claramente visibles.** Modelado de sistemas convencional, sin ser necesario el uso de métodos sofisticados.
- **Data Mining con modelo multivariable de la forma $y = f(x_1, x_2, \dots, x_n)$.** Aquí puede haber una gran relación entre la Lógica Borrosa y el *Data Mining* debido a que los objetivos de ambos mecanismos conducen a modelos que sean fácilmente comprensibles e interpretables. Por ello, el diseño de estos modelos está sustancialmente controlado por el Ingeniero de Conocimiento.
- **Modelos no funcionales.** El análisis de tablas relacionales no siempre implica la existencia de dependencias funcionales entre variables. Desde un punto de vista operativo, estos modelos pueden ser tratados como un conjunto de reglas con varias conclusiones: “*si condición entonces conclusión a ó conclusión b ó ...*”.

Como los patrones son inherentes al carácter relacional de los datos, cuando se asocian diferentes contextos a *clusters* similares, el modelado habitual es una mezcla de reglas y patrones. Es aquí donde la noción de *similaridad* adquiere especial relevancia. Hay muchos índices de similaridad definidos en la teoría de conjuntos borrosos, se suelen denotar con \equiv . $A = A' \approx 1$ denota que estas dos relaciones borrosas son muy similares. Los valores cercanos a 0 reflejan baja similaridad entre los objetos comparados. Como se podría esperar, la transición entre el modelo relacional y las dependencias funcionales no es abrupta. Eventualmente se pueden convertir patrones en relaciones funcionales admitiendo un mayor grado de incertidumbre en las funciones.

Si se tienen los *clusters* Ω_i y Π_j ($i = 1, 2, \dots, c; j = 1, 2, \dots, p$) inducidos respectivamente por los dos contextos A_1 y A_2 , se pueden calcular los índices de similaridad $[\Omega_i = \Pi_j]$ ($i = 1, 2, \dots, c; j = 1, 2, \dots, p$). La similaridad media se calcula mediante las similaridades entre los pares:

$$(\Omega = \Pi) = \frac{1}{c_1 c_2} \sum_{i=1}^{c_1} \sum_{j=1}^{c_2} (\Omega_i = \Pi_j)$$

Con esto se pueden dar cuatro situaciones:

$\Omega \equiv \Pi$ alta y $A_1 \equiv A_2$ baja. La naturaleza relacional de los patrones ha de ser retraída. El paso de los patrones a reglas tendría mucho error.

$\Omega \equiv \Pi$ baja y $A_1 \equiv A_2$ baja. Dos patrones distintos que podrían ser combinados con una función simple.

$\Omega \equiv \Pi$ alta y $A_1 \equiv A_2$ alta. Los patrones son altamente redundantes pero podrían ser combinados con una función simple.

$\Omega \equiv \Pi$ baja y $A_1 \equiv A_2$ alta. Dos patrones distintos. Es factible una función.

Cuanto menor sea el valor de la implicación $(\Omega \equiv \Pi) \rightarrow (A_1 \equiv A_2)$ (inducida por alguna t-norma), más fuerza tendrá la recomendación de no describir los patrones de forma funcional.

1.5.4. Diseño de modelos borrosos mediante Data Mining. El proceso DCPB.

En este se trabajo se proponen modificaciones del proceso KDD, tal como representa la fig. 4 que involucran incorporación de conocimiento en diferentes puntos, decisiones del usuario o experto y un resultado preparado para generar unos *prototipos conceptuales* (básados en la idea de Categorías Prototípicas Borrosas), que permitan evaluar situaciones nuevas a partir de dichos patrones, y establecer predicciones cuando estos patrones representan ciclos temporales. Las etapas del KDD modificado (a partir de ahora **Descubrimiento de Conocimiento Prototípico Borroso: DCPB**) son las mismas que las del KDD, pero teniendo en cuenta las siguientes modificaciones:

- El modelo borroso se desarrolla comenzando por el *Data Mining* de los datos disponibles. Este Data Mining no está guiado por los atributos, sino por heurísticas obtenidas en el preprocesado de los datos.
- Se definen una serie de contextos, denotados por A_1, \dots, A_n . Para cada contexto A_i , se ejecutan los algoritmos de *clustering* apropiados y se generan los

modelos basados en Categorías Prototípicas Borrosas sobre las regiones del espacio de entradas afines a los prototipos calculados por el método de *clustering*.

- El modelo completo se diseña considerando todos los contextos A_i , $i = 1, \dots, c$, y se asocia a cada uno de ellos una colección de *clusters* inducidos (ejecutando el esquema de *clustering*).

La arquitectura de este modelo es muy modular, porque se desarrolla sobre los *clusters* de los contextos individuales. Cambiando los contextos (bajo total control del Ingeniero de Conocimiento, según las heurísticas iniciales), se producen diferentes *clusters*, que pueden conducir a modelos significativamente distintos.

El aspecto más relevante de este modelo es su transparencia, tanto en su desarrollo como en su interpretación. La mayoría de los modelos borrosos se concentran en la capacidad de aproximación, sacrificando por tanto la capacidad de interpretación, construyendo modelos opacos o de “cajas negras” similares a los neuronales.

La deformación de los prototipos con afinidad a una situación real a evaluar se fundamenta en el criterio de estabilidad y convergencia del método. Al igual que en el entrenamiento de un perceptrón multicapa con el algoritmo de *backpropagation* se retropropaga el error del resultado esperado con respecto al realmente obtenido por la red, para ajustar la matriz de pesos y que cada iteración se obtengan mejores resultados, en el modelo propuesto se utiliza la afinidad (puede verse como diferencia o error) de la situación real con los prototipos de partida y trata de “subsanarse” esta disparidad modificando éstos prototipos para que se ajusten a la realidad de la forma más precisa posible. También influyen las ideas ya comentadas de que habitualmente los elementos menos representativos se suelen considerar similares a los más representativos y no al contrario, y de que los elementos que tienen mas atributos en común con los otros elementos de su categoría también son los que tienen menos en común con otras categorías.

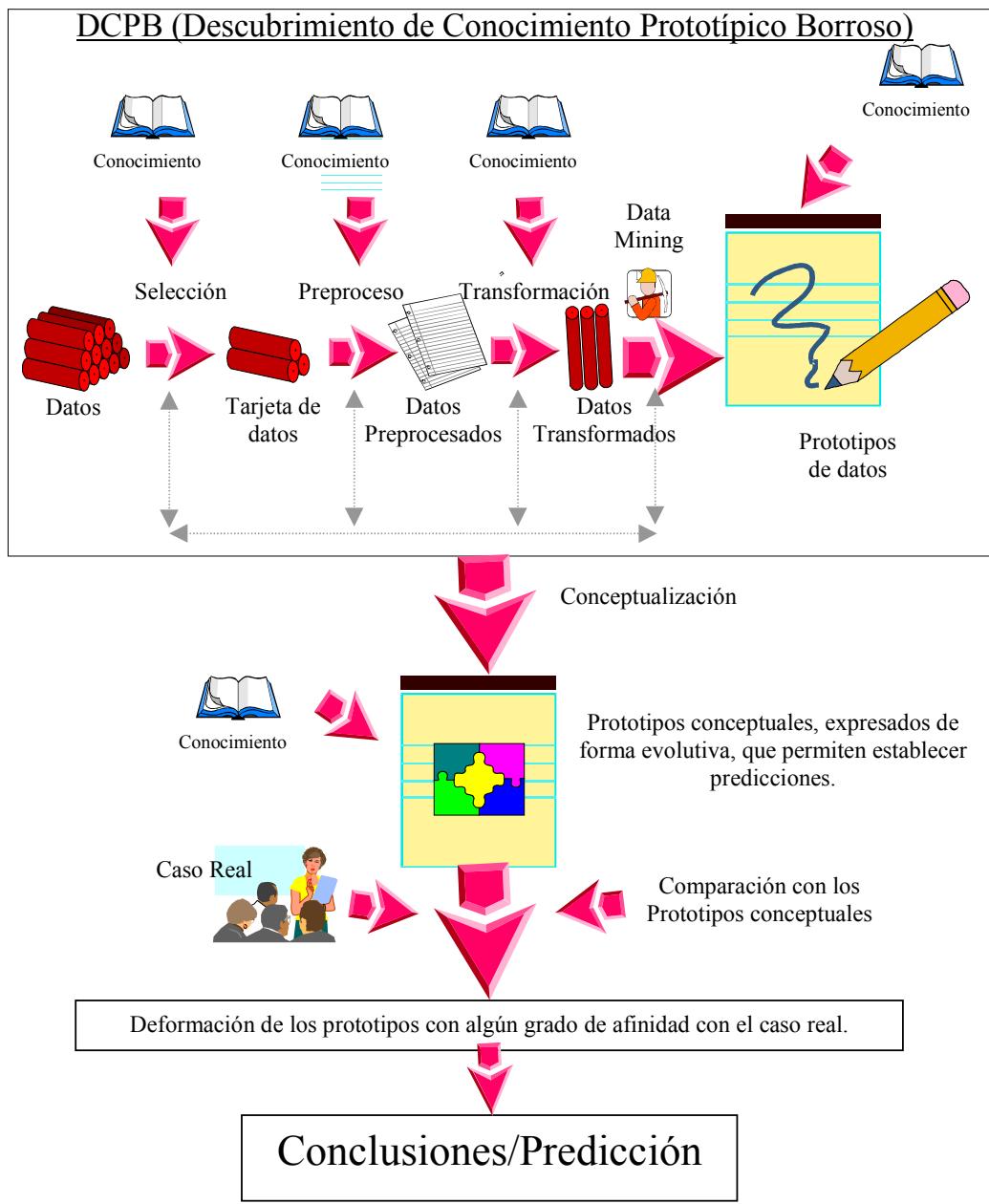


Figura 4: Proceso global de predicción incluyendo el DCPB.

1.5.5. Algoritmos para el proceso DCPB.

Un algoritmo de *Data Mining* dentro del proceso DCPB es una instancia de las componentes modelo/preferencia/búsqueda (ver Apéndice 1), por ejemplo un modelo de clasificación basado en una representación de árbol de decisión, un modelo de taxonomías basado en la verosimilitud de los datos usando heurísticas, etcétera. Habitualmente, estos algoritmos difieren mucho en cuanto a la representación y la función, pero no en los criterios de preferencia o búsqueda.

Cada problema real en particular requiere un algoritmo específico. Es tarea del Ingeniero del Conocimiento seleccionar las mejores opciones *a priori*, y si es necesario, revisar y probar modelos alternativos. A lo largo de este trabajo se muestran muchos casos que ilustran lo aquí expuesto.

2. Categorías Prototípicas Deformables Borrosas.

2.1. *Prototipos deformables.*

La noción de prototipo deformable surge en el ámbito del reconocimiento de patrones. En este contexto, las diferentes tareas en el reconocimiento de patrones suelen tener una estructura formal común: Primero hay un universo de objetos posibles U . Este universo se divide en clases de objetos que se agrupan bajo una etiqueta común (considérense, por ejemplo, los caracteres escritos). Para simplificar la tarea se define un cuadrado en el plano llamado *retícula*, y se asume que los caracteres a identificar aparecen individualmente en este cuadrado. En este caso, U se compone de todas las posibles configuraciones que pueden aparecer en la retícula, y si se incluyen los que no tienen sentido, U quedará constituido por todas las posibles configuraciones de dicha retícula. Lo que suele ocurrir, es que el número de objetos posibles es muy grande. Las posibles configuraciones de la retícula son infinitas, por lo que en la práctica, tanto U como la retícula se discretizan (fig. 5).

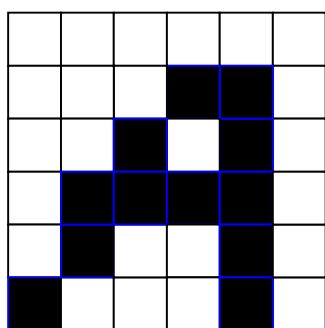
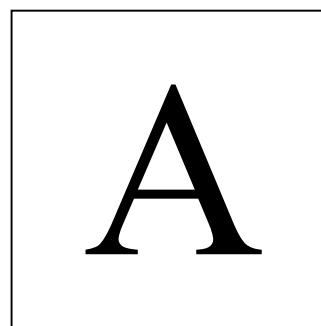


Fig. 5: Retícula discretizada.



Retícula.

2.1.1. Extracción de características.

Una *característica* es cualquier propiedad de un objeto patrón que puede ser medida. Ejemplos de características usadas frecuentemente en reconocimiento de patrones son:

1. Correlación del área ocupada por un objeto con un subconjunto dado de la retícula.
2. Máscaras: Funciones sobre la retícula que ponderan de forma diferente puntos diferentes.
3. Coeficientes de Fourier de desarrollos de Fourier bidimensionales de la función característica del subconjunto representado por el objeto y coeficientes de Fourier generalizados con respecto a varios sistemas de funciones ortogonales.
4. Momentos de un conjunto de puntos (normalmente el centroide).
5. Propiedades topológicas (líneas, esquinas ...).
6. Propiedades métricas (por ejemplo la relación entre el máximo y el mínimo diámetro de la envolvente convexa del objeto), etc.

Por lo tanto, la *extracción de características* consiste en mapear objetos patrón con puntos en R^n . Si las características han sido elegidas como relevantes para una clase patrón, entonces se mapeará entre la clase patrón y R^n . En este caso es cuando se denomina “*Cluster*”.

El *análisis cluster* es una técnica que determina la locación de *clusters* a partir de objetos simples conocidos (fase de aprendizaje) e identifica objetos desconocidos dependiendo del *cluster* al que están más cercanos. El proceso de aprendizaje puede ser interpretado como la estimación de la distribución de la probabilidad de los vectores característica en R^n , por lo que frecuentemente se utilizan técnicas estadísticas en el reconocimiento de patrones.

Resulta muy importante para el reconocimiento cómo separar varios *clusters* con superficies razonablemente simples en R^n . En el caso extremo, el separar los

conjuntos imagen de diferentes clases patrón en el espacio de características puede suponer un esfuerzo mayor que construir una lista explícita de todos los objetos del universo. Si el conjunto imagen de diferentes clases puede ser separado por hiperplanos lineales en R^n , las clases se denominan *linealmente separables*, pero esto es frecuente que no ocurra (Fig. 6).

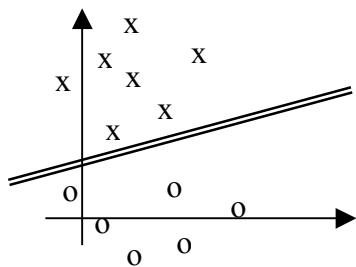
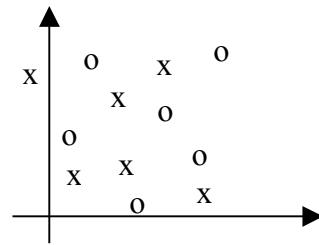


Fig. 6: Ejemplo de dos clases linealmente separables.



Ejemplo de dos clases linealmente inseparables

El mapeo de características en R^n , la determinación de los *clusters* y la separación de los *clusters* mediante hiperplanos se utilizan a menudo cuando las clases patrón no son estrictamente separables. Entonces, la estrategia consiste en posicionar las superficies de separación de forma que se minimicen las pérdidas, los elementos no clasificados. Este método suele ser utilizado cuando las características vienen dadas y no pueden ser definidas por el Ingeniero del Conocimiento. Por ejemplo, el diagnóstico médico podría ser visto como un reconocimiento de patrones, donde el universo U está constituido por los estados del organismo y algunos de esos estados son clasificados como “enfermedad”. El estado del organismo se mapea con síntomas observables (temperatura, análisis de sangre y orina, dolor, inflamación ...). A partir de estos síntomas, se trata de llegar a un diagnóstico que clasifica el estado del organismo como “saludable” o como poseedor de una (o varias) de una lista de enfermedades.

Los síntomas pueden ser considerados como un mapa de características, por lo que el diagnóstico automático intenta llegar a una clasificación correcta del vector de síntomas en el espacio de características. Los casos conocidos pueden ser utilizados para establecer *clusters* y definir las hipersuperficies de separación. Este método también puede ser interpretado de forma probabilística: Los casos de

entrenamiento son usados para estimar las distribuciones de probabilidad en el espacio de síntomas.

Debe resaltarse que habitualmente las clases de enfermedades son dadas más explícitamente que implícitamente; De hecho, el estado dinámico del organismo es en gran medida inobservable. Las clases patrón son definidas a través de ejemplos que son extrapolados (muchas veces se le llama al proceso de estimar *clusters* y superficies de separación “entrenamiento” o “aprendizaje”). Ocurre prácticamente lo mismo con el reconocimiento de voz y los caracteres escritos pero con una diferencia: Los sonidos y los caracteres son observables, a diferencia del estado del organismo y las características se pueden identificar, a diferencia de los síntomas, que vienen dados. Una vez elegidas las características, el tratamiento es el mismo: Algunos sonidos o caracteres son elegidos como conjunto de entrenamiento, y una vez que han sido establecidas categorías separadas en R^n , los sonidos o caracteres desconocidos pueden ser clasificados de acuerdo con dichas categorías.

Dicho esto, la elección de las características se revela como un aspecto clave para la consecución del éxito en un sistema de este tipo. Muchas de las características, de hecho, se eligen aleatoriamente, o en la base de una intuición vaga o incluso porque han llegado a ser populares en una profesión. Existe poca información que guíe en la elección de características en el proceso de Adquisición de Conocimiento. Posiblemente, el encontrar características perfectas sea equivalente a resolver el problema.

2.1.2. Prototipos Deformables e Implementación de Hodges.

En ese contexto se considera, por ejemplo, que los caracteres escritos a mano pueden ser vistos como imitaciones distorsionadas de prototipos ideales, sobre los cuales existe bastante consenso en cuanto a su representación. Así, un carácter cualquiera, puede ser básicamente considerado como compuesto de segmentos cuya posición se controla con una serie de parámetros.

Dado un carácter, se puede deformar su prototipo manipulando dichos parámetros hasta que coincide con el carácter en la mayor medida de lo posible (entre otras cosas, esto es debido a que la similaridad no tiene por que ser simétrica, se suele decir que un hijo se parece a un padre y no al revés), aunque es previsible la presencia de unos restos que no puedan ser cubiertos por las variaciones del prototipo. Estas discrepancias pueden ser caracterizadas por una función e_i de emparejamiento entre el objeto y el prototipo,

$$e_i(\text{objeto}, p_1, \dots, p_n) = | \text{objeto} - \text{prototipo}_i(p_1, \dots, p_n) |$$

donde

p_1, \dots, p_n	son los parámetros.
objeto	es el carácter a ser reconocido.
prototipo _i	es su prototipo.
$e_i(\text{objeto})$	es la función de emparejamiento.
$ $	es una medida (calculada y comprobada experimentalmente, por ejemplo las ya usadas para <i>clustering</i> o la norma euclídea) de la diferencia (similaridad) entre el objeto y el prototipo deformado.

y el e_i con el valor mínimo se corresponde con la mejor coincidencia entre el prototipo deformado y el objeto que se puede obtener.

La función e_i no es una buena medida de afinidad entre el objeto y el prototipo porque, por ejemplo, no tiene en cuenta la relevancia especial de determinados parámetros, no diferencia entre características esenciales y accidentales de objeto a examen o no tiene en cuenta lo que supone que un determinado parámetro sea muy dispar. A fin de salvar estas deficiencias, se introduce una función de distorsión $d(p_1, \dots, p_n)$, que relativiza el peso de los parámetros. Combinando las funciones e_i y d , se obtiene una función, llamada de afinidad que evalúa los efectos combinados. Esta nueva función de afinidad se puede definir como la combinación lineal de e_i y d ,

$$afin_i(\text{objeto}) = \min_{p_1 \dots p_n} (e_i(\text{objeto}, p_1 \dots p_n) + wd(p_1 \dots p_n))$$

donde,

w es una constante de pesos, y

mín denota el mínimo sobre todas las combinaciones posibles de valores de parámetros.

En este trabajo, el uso de la noción de prototipo deformable no se centra en el ámbito del reconocimiento de patrones, sino en la identificación de situaciones reales con categorías prototípicas borrosas que pueden ser parametrizadas [Olivas, 95b].

Se puede considerar la función de afinidad de un prototipo consigo mismo, ya que la coincidencia es exacta y no hay ninguna distorsión, como:

$$\text{afin}(\text{prototipo}) = 0$$

Cuando la coincidencia del “objeto” con el prototipo es buena (el nivel de distorsión es bajo) entonces *afin* (“objeto”) tendrá valores bajos (lo que contradice el uso habitual del concepto “afin”). Por ello, como *afin* por sí misma no es una función de pertenencia a un conjunto borroso, se puede definir la siguiente:

$$\mu(\text{"objeto"}) = 1 - \frac{\text{afin}(\text{"objeto"})}{\max(\text{afin})}$$

donde $\max(\text{afin})$ es el mínimo límite superior de *afin* en el mismo Universo U. Por lo tanto, μ (“objeto”) será 1 para el prototipo mismo, y más pequeña cuanto más diferentes sean los objetos del prototipo (podría haber otras funciones que cumplieran este cometido, pero en principio esta parece ser adecuada).

2.2. Categorías Prototípicas Deformables Borrosas.

Se definen, por tanto, las **Categorías Prototípicas Deformables Borrosas**, como una combinación lineal de Categorías Prototípicas Borrosas con la capacidad de adaptarse a una situación real, cuyos coeficientes son los grados de pertenencia a cada una de estas Categorías Prototípicas Borrosas.

Extendiendo la combinación lineal descrita en el concepto de *Prototipo Deformable*, al caso de un emparejamiento (afinidad) con más de una Categoría Prototípica Borrosa, la definición de una situación real será la siguiente:

$$C_{real}(w_1 \dots w_n) = |\sum \mu p_i(v_1 \dots v_n)|$$

Donde:

C_{real} Caso real.

$(w_1 \dots w_n)$ Parámetros que describen el caso real.

μp_i Grados de compatibilidad con las Categorías Prototípicas Borrosas distintos de 0.

$(v_1 \dots v_n)$ Parámetros de éstas Categorías Prototípicas Borrosas.

3. Evaluación de situaciones reales mediante razonamiento no-monótono.

Una vez obtenida la definición de una situación real mediante los mecanismos anteriormente descritos, ésta es representada como un marco (para una pequeña explicación del razonamiento no-monótono con marcos y negación por fallo, ver Apéndice 1), que servirá para evaluar las situaciones reales y generar las predicciones.

No obstante, parece indicado describir brevemente algunas funciones del ámbito del razonamiento borroso no-monótono, que no son habitualmente usadas en Sistemas Basados en el Conocimiento y que proporcionan un “*tuning*” que puede mejorar sensiblemente las prestaciones del sistema.

3.1. Razonamiento borroso no-monótono.

En el razonamiento borroso no-monótono, al igual que en la Ingeniería del Conocimiento (IC) en general, se produce un conflicto entre dos proposiciones (en este caso borrosas) cuando ocurre que siendo la primera proposición “X es A” y la segunda “Y es B” se cumple que $A \cap B = \Phi$. Se puede dar un caso particular, que

es que la segunda proposición sea “X es U”, siendo U el universo de discurso de X. En este caso a la segunda proposición se la denomina proposición *conflictiva*.

Se propone el uso de una función *media* m sobre el conjunto parcialmente ordenado de proposiciones borrosas (P_{\sim}, \leq) , de elementos de $(P_{\sim}, \leq) \times (P_{\sim}, \leq)$ en (P_{\sim}, \leq) , de forma que cumple que $A \cap B \leq m(A, B) \leq A \cup B$. La elección de esta función depende del tipo de conflicto, con lo que las consecuencias “X es A” e “Y es B” se cambian por $m(A, B)$, como proponen J. L. Castro y E. Trillas [Castro, 98]. Según estos autores, los patrones de reglas borrosas pueden clasificarse en los siguientes grupos:

- *Razonamiento por defecto* (reglas borrosas por defecto): Reglas donde la inferencia se hace sólo si no produce ninguna inconsistencia:

p₁: Si “X es A” entonces “Y es B”

p₂: “Y es U”

Un ejemplo puede ser “Si X es estudiante entonces X es joven”. Para cualquier estudiante se inferirá que es joven salvo que se sepa que no lo es. Para realizar la inferencia, se utiliza una extensión borrosa de una función *media* en la recta real, definida de la siguiente forma:

$$m(B_1, B_2) = \begin{cases} B_1 \cap B_2 & \text{si } B_1 \cap B_2 \neq \emptyset \\ B_2 & \text{en otro caso} \end{cases}$$

Siendo B₁ y B₂ las valoraciones de “X es A” e “Y es B” respectivamente.

- *Comportamiento prototípico*: Cuando no se sabe nada acerca del valor de una variable X, a menudo es razonable asumir que toma el valor borroso más común.

p₁: “X es A”

p₂: “X es U”

Por ejemplo, si no se sabe nada sobre el tamaño de una casa, se puede asumir que es de tamaño medio. Para realizar la inferencia, de nuevo se utiliza una

extensión borrosa de una función *media* en la recta real, definida de la siguiente forma:

$$m(B_1, B_2) = \begin{cases} B_1 & \text{si } B_2 = U \\ B_2 & \text{en otro caso} \end{cases}$$

Siendo B_1 y B_2 las valoraciones de “X es A” e “Y es U” respectivamente.

- *Asunciones de Mínimo Coste*: Cuando no se sabe nada acerca del valor de una variable X, y existe un coste asociado a algún posible valor borroso, a menudo se toma el valor borroso de coste mínimo.

p_1 : “X es A”

p_2 : “X es U”

Por ejemplo, si no se sabe cuan peligroso es un animal desconocido, usualmente se asumirá que es muy peligroso. Para realizar la inferencia se utiliza la misma extensión de la función media que en el caso anterior.

$$m(B_1, B_2) = \begin{cases} B_1 & \text{si } B_2 = U \\ B_2 & \text{en otro caso} \end{cases}$$

Siendo B_1 y B_2 las valoraciones de “X es A” e “Y es U” respectivamente.

- *Reglas borrosas complementarias*: Cuando se producen conflictos en el disparo de dos reglas borrosas de igual importancia y términos, resulta razonable aplicar una media entre ambas.

p_1 : Si “X es A_1 ” entonces “Y es B^1 ”

p_2 : Si “X es A_2 ” entonces “Y es B^2 ”

La media borrosa que se utiliza puede ser una de las tres siguientes:

1. *Media Aritmética Ponderada Generalizada*:

$$m(A, B)(x) = \sup_a A(a) \wedge B((x - (1-p)a)/p), p \in [0, 1]$$

2. *Media Geométrica Ponderada Generalizada*:

$$m(A, B)(x) = \sup_a A(a) \wedge B((x/a^{(1-p)})^{1-p}), p \in [0, 1]$$

3. *Media Aritmética Ponderada Borrosa:*

$$m(A, B)(x) = p A(x) + (1-p) B(x), p \in [0, 1]$$

Capítulo 3

V **iabilidad y Descripción del Sistema INCEND-IA.**

1. Viabilidad del sistema INCEND-IA.

Una vez definido el problema, la predicción de incendios, se plantea la siguiente pregunta: ¿Puede ser resuelto con un Sistema Basado en el Conocimiento?. Para responder a esta pregunta existen métodos sistemáticos y objetivos para calcular una medida de viabilidad del sistema. Se presenta un breve resumen del estudio de viabilidad llevado a cabo sobre el problema en estudio. Se ha utilizado el mecanismo propuesto en la metodología IDEAL, descrito en [Gómez, 97], pero solamente se hace una somera descripción del proceso, ya que no parece adecuado describirlo con más profundidad en este trabajo.

Los creadores de Sistemas Expertos no tienen fácil la descripción de las características que hacen que un problema sea adecuado para resolverlo creando un sistema de este tipo. Poder definir el problema es lo primero que se plantea cualquier organización que esté pensando en desarrollar un Sistema Basado en el Conocimiento. Dicho de otra forma, se plantean si el enfoque de un Sistema Experto resolverá eficientemente un problema particular.

Aunque no es posible ofrecer una respuesta simple y general a esta pregunta, es posible dar unas pautas de actuación que se han mostrado admisibles en la práctica. Por ello la tecnología de los Sistemas Expertos debe usarse si el desarrollo de un Sistema Experto es plausible, está justificado, es adecuado y si se va a tener éxito en su construcción. A continuación se considera cada apartado por separado:

1.1. Plausibilidad:

El objetivo de esta prueba es ver si es posible llevar a cabo el desarrollo del Sistema Experto. Para ello se enuncia la regla de la plausibilidad:

“Si existen verdaderos y genuinos expertos, y pueden proporcionar la mayoría de la experiencia necesaria, y son cooperativos, y son capaces de articular sus métodos y procedimientos de trabajo, y se dispone de suficientes casos de prueba, y la tarea no es muy difícil, y está adecuadamente estructurada, y no requiere sentido común, y solo exige habilidad cognoscitiva ... Entonces, es posible el desarrollo y construcción del Sistema Experto”.

1.1.1. El Desarrollo del Sistema INCEND-IA es Posible.

- Existen varios y genuinos expertos y son cooperativos:

Todo el Servicio de Defensa contra Incendios Forestales de la Xunta de Galicia, además de especialistas en diferentes aspectos específicos son accesibles y están dispuestos a colaborar.

- Los expertos concuerdan en las soluciones:

De un modo general se trata de unificar criterios para abordar el problema de la lucha contra los incendios forestales.

- Los expertos pueden articular sus métodos y procedimientos de trabajo:

Los procedimientos de trabajo son bastante estándar y pueden articularse.

- La tarea requiere sólo habilidad cognoscitiva.

- La tarea no es demasiado difícil:

Aunque muy compleja, la dificultad estriba en ser capaz de manipular un gran número de datos y variables en su conjunto.

- La tarea está bien estructurada y se entiende bien:

Después de un análisis meticuloso se puede estructurar perfectamente la tarea, y se pueden comprender perfectamente sus fases y estados.

- La tarea no requiere sentido común:

No se precisa un razonamiento de sentido común puro, sino un razonamiento rápido y exitoso basado en la formación técnica y la experiencia.

- Se dispone de suficientes y relevantes casos de prueba

Se dispone de todas las ocurrencias de incendios forestales de los últimos años y de la posibilidad de efectuar pruebas en paralelo.

1.2. *Justificación:*

El hecho de que sea posible desarrollar un Sistema Experto para una tarea particular no significa que esté justificado hacerlo. Para justificar el esfuerzo de un desarrollo de un Sistema Experto se debería seguir la siguiente regla:

“Si se necesita experiencia que resuelve una tarea útil y necesaria en entornos hostiles, penosos o poco gratificantes; o hay escasez de experiencia humana; o esa experiencia es necesaria simultáneamente en distintos y distantes lugares; o hay pérdida de dicha experiencia; o se espera una alta tasa de recuperación de la inversión; o no existen soluciones alternativas de tal modo que, por ejemplo, el enfoque de la programación convencional no es posible o satisfactorio ...

Entonces, la construcción del Sistema Experto está justificada”.

1.2.1. *El Desarrollo del Sistema INCEND-IA está Justificado.*

- Necesidad de experiencia en entornos hostiles, penosos o poco gratificantes:

Por definición un incendio forestal es hostil, penoso y poco gratificante. Además la toma de decisiones adecuada puede suponer evitar la pérdida de vidas humanas y de riquezas económicas y ecológicas.

- Escasez de experiencia humana:

Hay pocos expertos capaces de coordinar efectivamente este problema tan complejo, de hecho, muchos desastres que acontecen en España todos los años se deben a la falta de una buena toma de decisiones.

- Necesidad de experiencia en diversos lugares simultáneamente:

En todos los centros de coordinación se requiere disponer de los mejores medios para una efectiva toma de decisiones.

- Pérdida de la experiencia humana:

Algunos Agentes forestales, ingenieros y otros especialistas dejan el servicio por diferentes motivos, la experiencia que han acumulado durante años no debe perderse.

- Alta tasa de recuperación de la inversión:

Una buena gestión de los recursos disponibles en la lucha contra los incendios forestales puede provocar un gran ahorro económico (por ejemplo la hora de vuelo de un Helicóptero de vigilancia puede costar 150.000 pesetas).

- No existen soluciones alternativas:

No se conoce ningún sistema comercial que sea mínimamente parecido a éste.

- El enfoque de la programación convencional no es posible o satisfactorio:

Hasta el momento, ningún sistema convencional (gestor de bases de datos, extrapolaciones estadísticas...) ha logrado los resultados que se esperan con INCEND-IA.

1.3. Adecuación:

Sería inadecuado el que porque los Sistemas Expertos puedan ser útiles en determinadas circunstancias, el pensar en desarrollar todo mediante dicha tecnología. Es decir no basta con que el desarrollo de los Sistemas Expertos sea posible e incluso esté justificado para construirlo, pues además de las dos anteriores condiciones, también el Sistema Experto debe ser adecuado. Para determinar los factores clave que establecen cuando es apropiado desarrollar un Sistema Experto se va a seguir la siguiente regla.

“Si la naturaleza de la tarea está orientada a metas, es lo más reversible posible y afecta lo mínimo indispensable a la organización y es tal que, básica y primariamente, requiere manipulación y razonamiento simbólico que implica factores subjetivos y soluciones heurísticas; y la complejidad de la tarea no es

elemental pero es de conocimiento intensivo propio del dominio; y, por su alcance, es o puede reducirse a un tamaño manejable; y tiene un valor práctico; y necesita investigación básica para encontrar soluciones; y la transferencia de experiencia entre humanos es factible; y el Sistema Experto sirve a necesidades a largo plazo; y la experiencia está pobremente organizada; y no requiere respuestas en tiempo real inmediato ... Entonces, la construcción del Sistema Experto es adecuada”.

1.3.1. El Desarrollo del Sistema INCEND-IA es Adecuado.

Por su Naturaleza:

- Sirve a necesidades a largo plazo:

No es previsible que a corto plazo se termine el problema de los incendios forestales en España.

- La tarea requiere investigación básica:

Es necesario investigar técnicas básicas para atacar el problema. En este caso las aportaciones teóricas ya descritas en este trabajo.

- La tarea requiere manipulación simbólica:

Es muy útil el uso de lenguaje natural (etiquetas lingüísticas), y otras muchas variables y datos de todo tipo.

- La tarea requiere soluciones heurísticas:

Las heurísticas generadas a partir de la experiencia acumulada son la esencia del modo de abordar el problema de INCEND-IA.

Por su Complejidad:

- La tarea no es elemental:

Es una tarea con un alto nivel de complejidad.

- La cantidad de conocimiento exige una base de conocimiento.

- Los efectos de su introducción pueden planificarse:

Incluso es posible evaluar los beneficios que podría causar la implantación del sistema.

Por su Alcance:

- La tarea tiene valor práctico.
- La tarea es de tamaño manejable.
- Sirve a necesidades a largo plazo:

Los incendios forestales son una lacra que se prevé irá en aumento en los próximos años.

- La experiencia está pobremente organizada:

En cada centro de coordinación hay un nivel diferente de experiencia y muy pocas acciones se sistematizan.

- La tarea no requiere respuestas en tiempo real:

Aunque a veces las decisiones tienen bastante urgencia, en general se dispone de tiempo suficiente para meditar las decisiones. En todo caso, nunca es necesario lo que en sentido estricto se entiende por “tiempo real” en informática.

1.4. *Exito:*

Los Sistemas Expertos no son la panacea para alcanzar lo imposible, ni siquiera lo muy difícil. Una mera disposición para tener un Sistema Experto no garantiza que pueda construirse. Identificar una necesidad, no es suficiente para determinar una tarea apropiada para realizarla usando un Sistema Experto. Además de las consideraciones puramente técnicas para la aplicación de un Sistema Experto en la solución a un problema, existen otras muchas cuestiones, que de no tenerlas en cuenta pueden dar al traste con el uso de esta tecnología. Entre estas cuestiones cabe destacar las siguientes:

“Si los directivos están mentalizados de la importancia y eficiencia de la tecnología, y, por lo tanto, tienen expectativas realistas; y apoyan decididamente

el proyecto; y hay cambios mínimos en los procedimientos habituales y estándar; y los usuarios finales aceptan de buen grado, o por lo menos no rechazan “a priori” la introducción de un Sistema Experto; y los resultados no están muy influenciados por los cambios y vaivenes políticos; y el dominio es lo suficientemente estable; y las metas del proyecto global están adecuadamente integradas; y se efectúa una adecuada transferencia de tecnología y la tarea es de I+D o de carácter práctico; y se consigue una interfaz amistosa y explicativa; y los expertos concuerdan en sus soluciones ... Entonces el éxito del proyecto está garantizado”.

1.4.1. El Éxito del Sistema INCEND-IA está asegurado.

- Se efectúa una adecuada transferencia de tecnología:

Técnicas y herramientas como la Lógica Borrosa, algoritmos de aprendizaje, modelos de predicción, etcétera, se transfieren al sistema obteniendo resultados satisfactorios.

- Los directivos están mentalizados y tienen expectativas realistas:

En el ámbito de los incendios forestales se reconocen y utilizan las nuevas tecnologías.

- Hay cambios mínimos en los procedimientos habituales:

La implantación de INCEND-IA en los centros de coordinación no provoca cambios en el modo habitual de trabajo.

- Los usuarios finales no rechazan de plano la tecnología de los Sistemas Expertos:

Se ve simplemente como un Sistema útil para la toma de decisiones.

- Los resultados no están muy influenciados por los vaivenes políticos:

Aunque las diferentes administraciones suelen variar su estrategia de lucha contra incendios, el dominio y prestaciones de INCEND-IA se mueven a un nivel en que los cambios políticos no llegan a afectar.

- La tarea es de I + D y de carácter práctico.
- El dominio es suficientemente estable:

Desde los orígenes del hombre este dominio es estable. Solamente habrá cambios con la aparición de nuevas tecnologías.

- Las metas están adecuadamente integradas.

De lo aquí expuesto se extrae como conclusión que es viable la construcción del sistema INCEND-IA.

2. Descripción del sistema INCEND-IA.

Los usos de la informática en cuanto al tratamiento de información sobre incendios forestales son muchos, desde los procesadores de textos para construir “partes de incendio”, pasando por el mantenimiento de información histórica de diversa índole (tradicionalmente, el empleo de métodos informatizados, generalmente bases de datos, permite el almacenamiento sistemático de información relativa al problema), hasta la obtención de resultados en forma de estadísticas (incendios por día...), ver Apéndice 2.

Se ha pretendido construir este sistema para su utilidad global, por ello se concibe como un Sistema de Información Geográfica (en adelante SIG o GIS). Un SIG se puede considerar como un conjunto de mapas de la misma porción del territorio, donde un lugar concreto se puede localizar de alguna forma en todos los mapas incluidos en el sistema. El empleo de herramientas de información geográfica con capacidades de predicción a medio y largo plazo permitirían un análisis que es de gran ayuda en la toma de decisiones. INCEND-IA contiene todos los elementos necesarios para que pueda ser calificado como GIS (base de datos, bases de conocimiento para analizar y manipular los datos, y un sistema de interacción con el usuario) además de aportar un componente importante de Sistema de Ayuda a la Decisión.

Para desarrollar estrategias efectivas es necesario tanto el disponer de una base de datos perfectamente organizada, como de una herramienta que permita una eficaz manipulación de datos y un análisis espacial. Otros aspectos adicionales también deben ser cubiertos, como pueden ser la utilización del sistema informático por usuarios no cualificados, al menos en lo que se refiere a la entrada y consulta de datos. El sistema informático es adecuado cuando su implantación no supone un esfuerzo adicional en el trabajo cotidiano.

Durante los últimos 10 años, los sistemas de información geográfica (GIS) han aparecido como una prometedora herramienta para el análisis de las diferentes alternativas de distribución de recursos; además de proporcionar la posibilidad de recopilar y mostrar en pantalla diferentes recursos existentes.

Formalmente, se define un GIS como un sistema informático de hardware, software, datos geográficos espaciales, información descriptiva de los elementos geográficos para la consulta, mantenimiento, análisis y edición de información geográficamente referenciada.

Un SIG está constituido por objetos que pueden ser representados geográfica o espacialmente, es decir, pueden ser situados en un plano. La naturaleza de dichos objetos es diversa: suele estar asociada a algo físico, cultural, económico, etc. Símbolos, colores, tipos de líneas, son utilizados para representar las diferentes características espaciales en un mapa bidimensional. La información, por el contrario, no tiene porqué ser espacial. Cualquier base de datos, sea cual sea su tamaño y tecnología es en principio apta para trabajar con él.

Se conseguirá construir un auténtico sistema SIG cuando asociada la información con su componente geográfico se logren realizar operaciones entre los diferentes elementos y sus características temáticas, acompañado todo por poderosas herramientas de análisis.

Un SIG no suele almacenar información, sino que debe complementarse con máquinas de bases de datos externas que pueden ser desde grandes servidores de datos (Informix, Oracle, DB2, SQL Server) hasta pequeños sistemas basados en

MS Access o simples ficheros de DBase. En el caso de INCEND-IA los datos están contenidos en varias bases de datos construidas mediante MS Acces'97.

Para empezar, a riesgo de simplificar en exceso, las entidades que comprenderán la base de datos de INCENDIA son tan sólo dos: incendios y recursos. En recursos se tendrá a todo tipo de medios de diversa naturaleza contratados para la campaña, mientras que en los incendios se almacenarán todos los datos relevantes que se refieran a los sucesos que ocurran, en principio serían el objetivo del análisis principal. Estas entidades principales estarían rodeadas de muchas otras como situación de bases estables, tomas de agua, que incluirían información asociada, aunque sin influencia, salvo casos concretos, en el esquema organizativo.

2.1. Cartografía.

Cada central de defensa contra incendios es responsable de una comarca que tiene unos límites perfectamente definidos sobre los cuales ejerce sus funciones de control. Por tanto, será necesario que dicha área se encuentre totalmente digitalizada siguiendo la topología del GIS. Los mapas están divididos en cuadrículas que se corresponden con un Kilómetro cuadrado de superficie real y una representación analógica de las principales vías de comunicación, núcleos y masas forestales importantes, así como las de las zonas limítrofes.

Se utilizan representaciones analógicas de los Mapas militares usados por los propios servicios de Defensa contra Incendios, para poder estandarizar, por ejemplo, las referencias que se dan de un punto cuando tiene que actuar un medio aéreo.

2.2. La Aplicación.

El Sistema INCENDIA es una aplicación Windows construida bajo VISUAL C++ 5. Contiene todos los componentes propios de una aplicación Windows construida mediante las MFC (*Microsoft Foundation Classes*):

- Un armazón de aplicación,
- Arquitectura Documento/Vista,

- Interfaz de documentos múltiples (MDI),
- Una estructura de objetos interna representando a las entidades relevantes en el sistema (comarcas, municipios, incendios, recursos...).
- También cuenta con aspectos avanzados como acceso a las bases de datos de información del sistema mediante las técnicas DAO y ODBC y la utilización de controles *ActiveX* para conseguir una mejor presentación y un manejo más sencillo.

Como ya se ha indicado, el sistema informático será adecuado cuando su implantación no suponga un esfuerzo adicional en el trabajo cotidiano, evidentemente, tomando como referencia el trabajo que había que realizar antes de que el nuevo sistema comenzase a funcionar. La georreferenciación típica supone el indicar coordenadas geográficas mediante las pulsaciones del ratón sobre el mapa. En un trabajo futuro y para hacer más amplio el manejo se podría implementar la georreferenciación automática basada en nombres de zonas, y un sistema de autoaprendizaje por parte del sistema de los nombres de los lugares más relevantes...

Como se ha visto en esta introducción, la informática puede proporcionar un sistema sencillo (de cara al usuario) donde se pueden controlar y analizar los incendios que ocurran y la actuación de los recursos y acceder fácilmente a resultados sobre el análisis de la evolución temporal de los sucesos y así realizar una predicción de la evolución de la siniestralidad. Todo este componente irá acompañado de aplicaciones de informática de gestión que permitan informatizar otros aspectos de la actividad de defensa contra incendios, aplicaciones que entran más en el ámbito de la informática de gestión dada su naturaleza. como por ejemplo, tratamiento y corrección de partes, definición de fichas de trabajo...

2.3. *Bases de datos.*

En este sistema se parte de un modelo en el cual los objetos espaciales se representan de modo explícito llevando asociado un conjunto de aspectos temáticos. Existen varias posibilidades para organizar esta doble base de datos

(espacial y temática), en INCEND-IA se ha optado por almacenar en una base de datos relacional todos los datos referentes a las variables temáticas y en una “base de datos interna” organizada mediante una estructura de Objetos con matrices ráster relacionadas se maneja toda la estructura espacial. De esta forma se tiene el propio sistema INCEND-IA que maneja los mapas y todo lo referente a la localización de elementos geográficos y una base de datos que puede ser gestionada mediante MS Access '97 que maneja los atributos o variables temáticas. La clave del funcionamiento del sistema reside en la estrecha interrelación entre los dos componentes, establecida a través del identificador unívoco (clave primaria) de cada objeto geográfico que aparece en las dos bases de datos (la temática y la espacial), de manera que los cambios en un aspecto (por ejemplo el espacial) repercutan de modo inmediato en el otro (temático).

Al tratarse de una Base de Datos relacional los datos se organizarán en tablas que a su vez estarán relacionadas entre ellas.

Tabla FUEGOS.

Después de haber tratado la base de datos original de los datos de los incendios de la comarca de Lugo en 1991 y 1992 y haber detectado todas las inconsistencias y problemas que se han generado, se ha decidido construir una base de datos para poder automatizar el proceso de DCPB (Descubrimiento de Conocimiento Prototípico Borroso) necesario para la actualización de la predicción sin tener que realizar procesos previos de limpieza y eliminación de valores perdidos.

Campo	Tipo	Descripción
IDENT	Texto	Identificador del incendio
MUNICIPIO	Texto	Municipio donde se inicia el incendio
PARROQUIA	Texto	Parroquia donde se inicia el incendio
FECHA_INIT	Fecha/Hora	fecha en la que comenzó
HORA_INIT	Fecha/Hora	hora aproximada en la que comenzó
FECHA_CONT	Fecha/Hora	fecha en la que queda controlado el incendio
HORA_CONT	Fecha/Hora	hora en la que queda controlado el incendio
FECHA_FIN	Fecha/Hora	fecha en la que se sofocó totalmente
HORA_FIN	Fecha/Hora	hora aproximada en la que se sofocó
CAUSA	Texto	causa por la que se ha producido el incendio
SUP_TOTAL	Real	hectáreas totales arrasadas
SUP_ARBO	Real	hectáreas de bosque quemado
SUP_MONTE	Real	hectáreas de monte quemado

SUP_RASA	Real	hectáreas de pasto quemado
SUP_CULT	Real	hectáreas de cultivo quemado
TECNICOS	Entero	Personal técnico presente en el incendio
AUT_CIVIL	Entero	Autoridades Civiles presentes
AGEN_FOR	Entero	Agentes forestales presentes
PERSONAS	Entero	Personal de extinción (cuadrillas) que actuaron
MEDIOS	Entero	Medios mecánicos que actuaron en la extinción
OBSERV	Memo	Observaciones

Tabla 2: FUEGO.

Tablas de información de la comarca.

Mantienen datos de localización de elementos relevantes para la lucha contra incendios como es el caso de bases estables de recursos, tomas de agua...

Tomas de agua.

Campo	Tipo	Descripción
IDENTIFICADOR	Texto	identificador de la toma de agua
MUNICIPIO	Texto	municipio en el que se encuentra
COORDENADAX	Entero	coordenada X <i>ráster</i> de la toma de agua
COORDENADAY	Entero	coordenada Y <i>ráster</i> de la toma de agua
TIPORECURSO	Entero	tipo de recurso a los que puede abastecer
NOTAS	Memo	otros datos

Tabla 3: TOMAS DE AGUA.

Bases Estables

Campo	Tipo	Descripción
Base	Texto	identificador de la base
Municipio	Texto	zona en la que se encuentra
CoordenadaX Base	Entero	coordenada X <i>ráster</i> de la base
CoordenadaY Base	Entero	coordenada Y <i>ráster</i> de la base
Helipuerto	Booleano	tiene helipuerto (Si/No)
PistaAterrizaje	Booleano	tiene pista aterrizaje (Si/No)

Tabla 4: BASES ESTABLES.

Tabla PUNTOS SIGNIFICATIVOS.

Esta tabla se utiliza para definir zonas de alto riesgo para la defensa contra incendios por diversas causas como valor ecológico, nudo de comunicaciones. Será muy utilizada en las partes de consejo del sistema (Redistribución de recursos).

Campo	Tipo	Descripción
IDENTIFICADOR	Texto	Identificador del punto significativo
NOMBRE	Texto	nombre del punto significativo
MUNICIPIO	Texto	municipio en la que se encuentra
COORDENADAX	Entero	coordenada X del punto significativo
COORDENADAY	Entero	coordenada Y del punto significativo
RELEVANCIA	Real	relevancia[0,1]

Tabla 5: PUNTOS SIGNIFICATIVOS.

Tabla RECURSOS EN INCENDIOS.

Debido al cambio de diseño de la tabla FUEGOS se desconoce qué recursos han intervenido en ellos, para poder conocerlo se establece otra tabla de trabajo que es fruto de la relación entre las claves primarias de las tablas FUEGOS y RECURSOS. Indicándose en cada entrada el identificador del fuego y el identificador del recurso que actuó. Esta tabla se completa automáticamente a través del programa principal (salvo que exista alguna equivocación al respecto).

El sistema Informático permite el seguimiento y control de la situación en que se encuentran los medios humanos, los medios aéreos, los vehículos de extinción y la maquinaria pesada que interviene en un incendio.

Tabla MEDIOS HUMANOS.

En la base de datos se tiene todo el personal a lo largo de la campaña: retenes de especialistas, brigadas especiales, técnicos, agentes forestales, policía autonómica y grupos de voluntarios, entre otros. Para cada uno de estos grupos se controlan los siguientes datos:

- Situación operativa del personal: Activo, en estado de espera para actuar, descanso, fuera de servicio.
- Condiciones de avituallamiento: Si han comido o no, fecha y hora del ultimo avituallamiento.
- Cumplimiento de horarios: Hora en que comenzó a trabajar, hora en que dejó de trabajar, tiempo de trabajo acumulado, tiempo de descanso acumulado.

Tabla MEDIOS AÉREOS.

En este caso se tienen en la base de datos todos los medios aéreos contratados para la campaña de incendios. Para cada uno de estos aparatos se consideran las situaciones siguientes:

- Estado operativo en que se encuentra: Activo, operativo pero sin actuar, emplazado, repostando, fuera de servicio.
- Condiciones de la intervención en los incendios: Hora de comienzo de vuelo, hora de final del vuelo, tiempo acumulado de vuelo, hora a la que se retira para repostar.

Tabla VEHÍCULOS TERRESTRES.

Para la extinción se dispone de un parque de vehículos formado por autobombas, nodrizas, y de patrullaje provistas de depósitos de agua y las unidades móviles ya mencionadas.

Además en los incendios de importancia pueden intervenir maquinaria pesada, básicamente tractores *bulldozer* o tipo *skider* con pala frontal y *ripper* para la realización de cortafuegos y líneas de defensa.

Las situaciones a controlar para estos vehículos son las siguientes:

- El estado en que se encuentra la máquina: Activa, operativa pero sin actuar, fuera de servicio.
- Condiciones de intervención: Hora de comienzo, hora de finalización, tiempo acumulado de trabajo.

2.4. *Cartografía.*

Unidad mínima cartografiable.

En un sistema ráster un elemento esencial es el tamaño del elemento base de la rejilla (unidad mínima cartografiable) y, asociado con él, el número total de filas y columnas de la cuadrícula.

Para determinar la unidad mínima cartografiable se debe tener en cuenta el tamaño del elemento base necesario para incluir en la base de datos los hechos de realidad de menor tamaño. Tomando ejemplos de otros sistemas que gestionan incendios la región controlada suele ser de 10 km^2 , pero esto es demasiado grande para la comunidad Gallega ya que en esos 10 km^2 pueden existir multitud de montes, lugares donde ubicar recursos, incendios en activo... Para conseguir un equilibrio entre la representatividad y la tratabilidad de los elementos georreferenciables se opta por una rejilla de 1 km^2 que se supone significativa para todos los elementos a tratar.

En el mapa de un municipio esta rejilla se transforma en un cuadrado de tamaño $16 \times 16 \text{ pixels}$. Las matrices representativas de los mapas del SIG se representarán por enumeración exhaustiva. (27×24 en mapas municipales).

Entrada de Datos y digitalización ráster.

La fuente utilizada para crear este sistema ráster es la cartografía analógica. A partir de ahí se ha utilizado principalmente la digitalización ráster manual y la utilización ocasional de otras técnicas de rasterización.

Digitalización ráster manual.

El proceso de digitalización manual se ha realizado superponiendo sobre el mapa analógico que sirve de fuente (mapa militar), una cuadrícula cuya celda base tiene un tamaño que es precisamente el de la unidad mínima cartografiable. A continuación se ha codificado cada uno de los *pixels* con un número que indicará el carácter de la información analógica que está contenida en él, siguiendo distintos tipos de codificación dependiendo del tipo de objeto espacial de la realidad que se necesite representar.

Los resultados del proceso de digitalización se almacenan en ficheros *bitmap* (.bmp) para que puedan ser tratados por la aplicación:



Figura 7: Mapa del ayuntamiento de Lugo.



Figura 8: Mapa del ayuntamiento de Palas de Rey.

2.5. Representación de objetos puntuales.

- Mapas municipales: Se opta por una codificación de presencia/ausencia. El resultado que se obtiene es un solo mapa binario con 1 en los *pixels* que representan a zonas del municipio y 0 en las zonas que pertenecen a otros municipios (aunque sean de la misma comarca).
- Recursos en un mapa: A cada uno de los *pixels* se asigna un valor cuantitativo que enumera la cantidad de recursos que están situados en esa zona. El resultado es la obtención de un mapa *ráster* cuantitativo.
- Incendios en un mapa: Se opta por representar un incendio en el mapa con el valor de un índice (valor temático) en una lista de incendios que se mantiene en paralelo, de esta forma se puede acceder rápidamente a sus características. De la misma forma se *rasterizan* los puntos significativos (de alto riesgo) dentro del mapa

Tanto los recursos como los incendios no son puramente una digitalización manual si no que se automatiza mediante el uso del sistema colocando estos elementos dentro del mapa del municipio o comarca que sí está digitalizado manualmente.

Representación de objetos lineales.

Las carreteras y caminos, principalmente en los mapas municipales se podrían codificar de forma presencia/ausencia dentro del mapa, pero es más significativo (sobre todo para la evaluación de proximidad de elementos) que se indiquen con el valor temático del elemento indicando la calidad de la carretera o del camino.

Objetos reales poligonales.

El Mapa general de la Comarca se puede representar de muchas maneras. Las fronteras entre los municipios dentro del mapa de la comarca: mediante presencia/ausencia de frontera, mediante el municipio dominante en la zona de frontera, el valor del municipio de dentro de la zona... Estos valores se introducirán de la misma forma que han sido construidos los mapas municipales de manera totalmente manual.

Otras técnicas.

La digitalización *ráster* automática se deja para la obtención de datos exteriores por medio de teledetección o mediante el análisis de datos. Se podría incluir aquí la codificación de los incendios y de los recursos que no es propiamente manual, pero no es una codificación automática sino una forma de hacer dinámica y asistida mediante el entorno informático de una codificación manual.

Los valores de las matrices con 1 son puntos dentro del municipio, los valores 0 son los puntos que pertenecen a municipios limítrofes de la propia comarca o de otras comarcas dentro o no de la provincia de Lugo.

Lugo

Palas de Rey

Figura 9: Ejemplos de Codificación Inicial de los mapas municipales de Lugo y Palas de Rey

3. La aplicación.

3.1. El armazón de la Aplicación.

Un armazón de aplicación es un “conjunto integrado de componentes de software orientado a objetos que ofrece todo lo que se necesita en una aplicación genérica”. También se puede ver un armazón de aplicación como un superconjunto de una biblioteca de clases. Una biblioteca de clases ordinaria es un conjunto aislado de clases diseñadas para ser incorporadas en cualquier programa. Pero un armazón de aplicación define la estructura del propio programa. El sistema que se presenta utiliza el armazón de aplicaciones por defecto que presentan todas las aplicaciones que se basan en diferentes versiones de las MFC (*Microsoft Foundation Classes*).

La utilización del armazón de aplicación proporcionará la estructura genérica de una aplicación Windows, para darle una funcionalidad propia, el desarrollador tendrá que introducir su propio código a la aplicación.

3.2. Arquitectura Documento-Vista.

La arquitectura documento-vista separa los datos de la visión que tiene el usuario de los mismos. De esta forma se pueden obtener múltiples vistas de los mismos datos.

En una aplicación de la biblioteca MFC, los documentos y vistas se representan mediante ejemplares de clases de C++. El código de la clase base del documento interactúa con los elementos de menú que abren los documentos; la clase documento derivada realiza las operaciones de lectura/escritura real de los datos del objeto documento. La clase de base de la vista representa una ventana que está contenida dentro de una ventana de marco; la clase de vista derivada interactúa con su clase de documento asociada y realiza la visualización de la aplicación. La clase de vista derivada y sus clases de base gestionan los mensajes de Windows. La biblioteca MFC dirige todas las interacciones entre los documentos, vistas, ventanas de marco y el objeto aplicación, principalmente a través de funciones virtuales.

En INCEND-IA el documento principal es la Comarca, en el caso particular que nos atañe la Comarca de Lugo. Para este documento único existen diferentes vistas:

- La vista de Comarca, que presenta el mapa completo de la zona y los datos estadísticos generales.
- La vista Municipal que presenta el mapa del municipio, la lista de recursos general y los diferentes elementos georreferenciados sobre el mapa. Se pueden ver todos los municipios de la comarca en diferentes ventanas.

3.3. El Interfaz de Documentos Múltiples.

La Interfaz de Documentos Múltiples (MDI) se diseñó para simplificar el intercambio de información entre documentos, todos bajo el mismo tejado. Con una aplicación MDI se pueden mantener abiertas varias ventanas, sin tener abiertas varias copias de la aplicación. El intercambio de datos es más fácil si se pueden ver y comparar varios documentos simultáneamente.

En una aplicación MDI se pueden poseer varias ventanas con un documento diferente cada una o cada ventana con una vista del documento distinta a las demás. La ventana principal nunca se duplica pero actúa como contenedor de las otras ventanas, y se le llama la ventana madre. Las ventanas en las que se muestran los documentos individuales se llaman ventanas hijas. Cuando la ventana madre cambia de posición en el escritorio, le siguen las hijas. Sin embargo, las ventanas hijas existen independiente de la ventana madre. Se pueden abrir y cerrar tantas ventanas hijas como se quiera, que podrán tener incluso diferentes funciones.

El armazón de aplicaciones permite utilizar múltiples ventanas hija tipo MDI, para representar varias vistas pudiéndose crear múltiples objetos tipo vista de la misma clase vista (la vista normal) y poder utilizar dos o más clases vista en la misma aplicación.

INCEND-IA es una aplicación MDI que maneja múltiples vistas de un solo documento (la Comarca). Las vistas posibles son la Comarca y el Municipio, pudiéndose consultar mediante varios objetos vista los diferentes municipios que pertenecen a la comarca. Todas las operaciones que se realicen sobre cualquiera de las vistas tienen efecto en el documento y pueden verse reflejadas en las demás vistas (del mismo o diferente tipo, del mismo o de distinto municipio).

3.4. Estructura de Objetos Interna.

Las clases de la Estructura de Objetos interna serán comentadas siguiendo el siguiente modelo de ficha. Las operaciones sobre los objetos son explicadas en la parte de Automatización de procesos.

<i>NOMBRE DE LA CLASE</i>
<i>Descripción</i>
<i>Atributos</i>
<i>Elementos de implementación</i>
<i>Constructores</i>

Tabla 6: Modelo de ficha para representar los objetos.

COMARCA
<i>Una comarca es un conjunto de municipios y de todos los elementos relevantes para representar todas las variables temáticas del problema de los Incendios Forestales.</i>
<ul style="list-style-type: none"> - mapa digital gráfico - matriz <i>ráster</i> básica - Lista de Municipios presentes en la comarca - Lista de Incendios activos en la comarca - Lista de Recursos destinados en la comarca - matrices <i>ráster</i> de información auxiliar
Dos listas de índices que indican los índices de los recursos de la comarca que están activos (en una de ellas) e inactivos (en la otra).
<p>El constructor de la comarca que toma los datos de diversas procedencias</p> <ul style="list-style-type: none"> - Recursos: Tabla de la BD destinada al efecto - Municipios: los construye a partir de datos manualmente en un fichero <i>ráster</i> con información de identificadores de municipio y de la matriz <i>ráster</i> de cada uno. - Mapa y matriz <i>ráster</i> vienen como recursos propios de la aplicación - matrices <i>ráster</i> de información : geocodificación de los datos procedentes de las diferentes tablas de la base de datos

Tabla 7: Objeto Comarca.

MUNICIPIO
<i>Un municipio viene a contener todos los elementos que le representan (datos geográficos)</i>
<ul style="list-style-type: none"> - Identificador único de cada Municipio - Una matriz <i>ráster</i> de presencia/ausencia que delimita sus fronteras en el mapa municipal - un Bitmap que represente el mapa municipal - Aspectos estadísticos como el número de incendios activos en él, y el de recursos destinados
Se construye un municipio a partir de un identificador, un bitmap con el mapa y la matriz <i>ráster</i> correspondiente a él

Tabla 8: Objeto Municipio.

FUEGO
<i>Un incendio viene determinado por las coordenadas donde se presenta, posee un estado y unos recursos asignados a él</i>
<ul style="list-style-type: none"> - Un identificador (sirve para identificar el objeto con el registro de la BD) - Estado del incendio (activo, controlado, ...) - Coordenadas municipales del incendio - Una lista con los recursos asignados a ese incendio
La lista de recursos asignados al incendio que se transforma en un vector (<i>array</i>) de índices de los recursos de la lista comarcal

Se construye un incendio a partir de un identificador, unas coordenadas municipales (coordenadas + municipio). Por defecto está el estado del incendio que se pasa a activo

Tabla 9: Objeto Fuego.

RECURSO
<i>La definición de un recurso depende de su estado, ya que si no está activo no serán válidas sus coordenadas. Para su identificación global posee un nombre y un tipo</i>
<ul style="list-style-type: none"> - Coordenadas Ráster Municipales (sólo para recursos activos Punto Posicion + IdMunicipio Municipio) - Identificador del recurso - Identificador del tipo de Recurso - Identificador del estado: ESTADO (asignado, activo o inactivo)
<i>Se construye un recurso primeramente con su identificador y su tipo, para luego según su situación llenar el resto de los campos</i>

Tabla 10: Objeto Recurso.

3.5. Acceso a Bases de Datos.

Visual C++ incluye todos los componentes previstos para escribir aplicaciones de bases de datos en C++ para Microsoft Windows incluyendo dos sistemas de acceso a bases de datos distintos: Conectividad abierta de bases de datos (ODBC), y objetos de acceso a datos (DAO).

El estándar ODBC.

El estándar de conectividad abierta de bases de datos (ODBC) de Microsoft no sólo define las reglas gramaticales SQL, sino también el interfaz de programación en lenguaje C/C++ para interactuar con una base de datos SQL. Así se hace posible que un único programa en C o C++ compilado acceda a cualquier DMS que disponga de un controlador ODBC. El equipo de herramientas de desarrollo software (SDK) ODBC, que se incluye con Visual C++, incorpora controladores

de 32 bits para archivos DBF, bases de datos Microsoft Access MDB y un largo etcétera.

Con las clases de la MFC de Windows se utilizan objetos C++ en vez de descriptores de ventanas y descriptores de contexto de dispositivo; con las clases ODBC de la MFC se utilizan objetos en lugar de descriptores de conexión y descriptores de sentencia. El descriptor de entorno se almacena en una variable global sin ser representando por ningún objeto C++. Las dos clases ODBC principales son *CDatabase* y *CRecordSet*. Los objetos de la clase *CDatabase* representan conexiones ODBC con fuentes de datos, mientras que los objetos de la clase *CRecordset* representan conjuntos de filas desplazables. La documentación de Visual C++ utiliza el término “conjunto de registros” en lugar de “conjunto de filas”, para ser consistente con Microsoft Visual Basic y Microsoft Access.

DAO, COM y el motor de base de datos Jet de Microsoft.

Una característica de DAO es un conjunto de interfaces COM, que, como todas las interfaces COM, no son más que especificaciones (conjuntos de declaraciones de función virtual pura). Estas interfaces tienen nombres como *DAOWorkspace*, *DAODatabase* y *DAORecordSet*.

La otra característica de DAO es la implementación de esas interfaces. Microsoft proporciona el módulo DAO350.DLL de COM, que se conecta a la misma DLL del motor de base de datos Jet que se utiliza en el producto de base de datos de Microsoft Access.

La biblioteca MFC tiene las cinco clases de base de datos DAO siguientes: *CDaoWorkspace*, *CDaoDatabase*, *CDaoRecordset*, *CdaoTableDef* y *CDaoQueryDef*. Estas clases envuelven más o menos a los interfaces COM de nombres similares (*CDaoRecordset* envuelve a *DAORecordset*, por ejemplo). La clase *CDaoWorkspace* realmente envuelve a dos interfaces, *DAOWorkspace* y *DAODBEngine*. El envoltorio de la MFC es bastante completo, por lo que sólo se necesita realizar llamadas directas al DAO de COM cuando se necesita acceder a ciertas características de seguridad de base de datos. Al utilizar la biblioteca MFC, ésta se ocupa de la gestión de contadores de referencia; si llama directamente a

DAO, debe asegurarse de llamar a *Release* para los interfaces. Los *Wizards* de Visual C++ (*AppWizard* y *ClassWizard*) admiten DAO totalmente. Se puede utilizar el *AppZiward* para generar una aplicación de acceso a base de datos pura y el *ClassWizard* para generar una clase específica de tabla que se deriva de *CDaoRecordSet*.

Los Gráficos y el interfaz GDI.

Cualquier SIG depende de las imágenes para poder ser útil, por lo que INCEND-IA utiliza la cartografía en ficheros de mapas de bits de Windows. Los objetos mapa de bits GDI están representados por la clase *CBitmap* de la Biblioteca MFC. El objeto mapa de bits GDI tiene asociada una estructura de datos de Windows mantenida por el módulo GDI de Windows que es dependiente del dispositivo.

3.6. Automatización de procesos.

Funciones del sistema.

- *Funciones de entrada de información:* Mediante el sistema INCEND-IA se pueden introducir todos los datos que provienen de las emisoras de radio de los elementos participantes en la lucha contra los incendios (centrales de radio, agentes forestales, brigadas, vehículos terrestres y aéreos...) con un formato que pueda ser manipulado y controlado de forma totalmente automática. Para ello se implementan mecanismos de entrada (localización de incendios y recursos, diálogos para consulta y modificación de datos...) y procedimientos que permiten eliminar errores, redundancias o pérdidas en la información que pueda llegar del mundo real. En este sistema el proceso de obtención de la información es principalmente dinámico ya que gestiona puntualmente la situación en un momento dado.
- *Salida, representación gráfica y cartográfica de la información:* Muy relacionados con los anteriores se utilizarán para mostrar al usuario todos los datos incorporados en la Base de Datos del SIG y el resultado de operaciones sobre ellos, permitiéndose la visualización de mapas, formularios de consulta de datos, tablas, informes estadísticos...

- *Las funciones de gestión de la información espacial:* Con la cuales extraen de los datos y posiciones que interesan en cada momento, y es posible reorganizar todos los elementos integrados en ella de diversas maneras (redistribuir recursos...).
- *Las funciones analíticas:* Son el elemento más característico de un SIG. Facilitan el procesamiento de los datos integrados en él de modo que sea posible obtener mayor información, y con ella mayor conocimiento del que inicialmente se disponía.

3.7. Operaciones en la aplicación

Se describirán las operaciones con el siguiente modelo de ficha de operación:

NOMBRE DE LA OPERACIÓN	TIPO
OBJETOS INVOLUCRADOS	
CONDICIONES DE ACTIVACIÓN	
EFFECTO	
PSEUDOCÓDIGO	

Tabla 11: Modelo de ficha de operación.

Inserción de un incendio	Entrada
Objetos involucrados: La clase Fuego y la Clase ListaFuegos. Afecta al objeto ListaFuegos del Objeto Comarca	
Condiciones de activación: El Modo de operación estará a ACCION y se activará con el evento WM_LBUTTONDOWN	
Efecto: Introduce un incendio en la comarca reseñándolo en el mapa, en la base de datos y en la lista de incendios de la estructura interna	

1. Si MODO_OPERACION no es ACCION entonces TERMINAR
Si no CONTINUAR
 2. Determinar si la posición donde se hizo *click* pertenece al mapa *ráster*
 - 2.1 Corregir el valor del punto de *pixels* a coordenadas *ráster*
 - 2.2 Obtener el Valor de la Matriz *Ráster*
 3. Según sea el valor regresado

INVALIDA: TERMINAR

FUERAMUNICIPIO: Presentar Mensaje al respecto

OTROSVALORES:

 - 3.1. Construir el incendio a almacenar
 - 3.1.1 Invocar al diálogo correspondiente
 - 3.1.1.1 Inicializar variables
 - 3.1.1.2 Cargar el diálogo
 - 3.1.1.3 Obtener las variables correspondientes del diálogo
 - 3.1.2 Obtener el valor de la posición donde se va a colocar
 - 3.1.3 Construir la cadena identificadora
 - 3.1.4 Invocar al constructor del objeto
 - 3.3. Almacenar los datos del incendio procedentes del diálogo en la BD
 - 3.2. Almacenar el fuego en la lista de fuegos de la Comarca
 - 3.3. Indicar su posición en la matriz *ráster*
4. Restaurar el modo de operación a INFORMACION

Tabla 12: Inserción de un incendio.

Extinción de un incendio	Entrada
<p>La clase Fuego y la Clase ListaFuegos. Afecta al objeto ListaFuegos del Objeto Comarca eliminando el Fuego correspondiente. También afecta a la lista de Recursos de la Comarca cambiando a estado ACTIVO los que estaban en estado ASIGNADO y en ese incendio.</p>	
<p>Es un efecto lateral de la operación de Consulta/Modificación de la información del incendio. Se efectúa cuando el estado se ha cambiado a Extinguido.</p> <ol style="list-style-type: none"> 1. Volcar la información del incendio en la BD y marcarla como cerrada 2. Eliminación del Incendio de la Estructura Interna <ol style="list-style-type: none"> 2.1. Liberar los recursos asignados (pasar su estado a ESTADOACTIVO y mantener su posición) 2.2. Eliminar el Incendio de la Lista 2.3. Actualizamos el valor de la matriz <i>ráster</i> (reflejar si quedan recursos ahí) 2.4. Repintar el mapa (sólo la región del incendio) 	

Tabla 13: Extinción de un incendio.

Asignación de un recurso a un incendio	Interna
El objeto en cuestión de la lista general de recurso y el objeto incendio al que ha sido asignado	
Internamente se realiza cuando se arrastra un recurso sobre una posición que está afectada por un incendio	

Cambia los datos de posición del recurso que hace coincidir con los del incendio, refleja que está en estado ASIGNADO e inserta su índice en el incendio.
<ol style="list-style-type: none"> 1. Recuperar el objeto recurso de la lista general de recursos <ol style="list-style-type: none"> 1.1 Recuperar su índice de los índices de inactivos 1.2 Recuperar el objeto de la lista a partir de su índice 2. Obtener el objeto incendio de la lista de incendios activos de la comarca 3. Modificar las características del recurso <ol style="list-style-type: none"> 3.1 Pasarlo a estado ESTADOASIGNADO 3.2 Hacer coincidir su nueva posición (municipio y coordenadas) con las del incendio 3.3 Indicar a que incendio de la lista está asignado el recurso 4. Introducir el recurso en el incendio

Tabla 14: Asignación de un recurso a un incendio.

Consulta/Modificación de información sobre el incendio	Entrada/Salida
	La clase Fuego y la Clase ListaFuegos. Afecta al objeto ListaFuegos del Objeto Comarca
	El Modo de operación estará en INFORMACION y se activará con el evento doble <i>click</i> del botón izquierdo del ratón
	Se presenta con un diálogo con la información relevante del incendio. Las modificaciones que se realicen sobre estos datos se verán reflejadas en la Base de Datos. El cambio de estado además se reflejarán en la estructura interna y se mostrará el incendio con un color correspondiente a él (si se pone a extinguido ver operación posterior), también se puede consultar y modificar la información de los recursos asignados a ese incendio (ver operación propia)

La fase de consulta se termina en el paso 3.2 luego se realizan las acciones propias a la modificación de la información

1. Si MODO_OPERACION no es ACCION entonces TERMINAR
Si no CONTINUAR
2. Determinar si la posición donde se hizo *click* pertenece al mapa *ráster*
 - 2.1 Corregir el valor del punto de *pixels* a coordenadas *ráster*
 - 2.2 Obtener el Valor de la Matriz *Ráster*
3. Según sea el valor regresado

INVALIDA: TERMINAR

FUERAMUNICIPIO: Presentar Mensaje al respecto

SI es un Valor indicador de Incendio ENTONCES

 - 3.1. Invocamos al diálogo de información
 - 3.1.1 Recuperar el incendio correspondiente de la EI
 - 3.1.2 Recuperar los datos del incendio de la Base de Datos
 - 3.1.3 Inicializar campos del diálogo con datos recuperados
 - 3.1.4 Cargar el diálogo
 - 3.2 OPERACIONES DEL USUARIO DENTRO DEL DIALOGO
 - 3.3 Obtener los cambios producidos en la información
 - 3.3.1 Volcar variables relevantes modificadas a BD
 - 3.3.2 Estudiar el cambio de estado
 - 3.3.2.1 Insertar fechas de cambio de estado en BD
 - 3.3.2.2 Cambiar el estado del incendio en la Estructura interna (Si el nuevo estado es EXTINGUIDO mirar operación propia)
 - 3.3.2.3 Repintar el mapa.

Tabla 15: Consulta/Modificación de información sobre el incendio.

Colocar/Recolocar un recurso en el mapa	Entrada
Las listas de recursos y los índices de control de la clase Comarca. Se modifican los objetos Recursos de la lista y los índices de Activos e inactivos.	
Consiste en arrastrar un recurso de cualquiera de las dos listas de recursos de la comarca (lista de recursos inactiva y activa) soltándolo en una posición de uno de un municipio. Eventos relacionados son BeginDrag de las dos listas WM_LBUTTONDOWN que da lugar a la ejecución de la tarea en sí	
Vuelca todos los datos del diálogo al la Base de Datos y marca estos datos como finalizados (no de trabajo inmediato), además elimina toda presencia del incendio en la EI (lista de incendio y matriz <i>ráster</i>) y deja allí los recursos que ha tenido asignados para que sean luego recolocados donde proceda.	

1. Si MODO_OPERACION no es ASIGNACION o No Hay Recurso A Mover entonces TERMINAR
Si no CONTINUAR
2. Determinar si la posición donde se hizo *click* pertenece al mapa *ráster*
 - 2.1 Corregir el valor del punto de *pixels* a coordenadas *ráster*
 - 2.2 Obtener el Valor de la Matriz *Ráster*
3. Si el recurso a Mover está ya activo
 - 3.1 Desactivar Recurso (ver operación propia)
 - 3.2 Actualización de los datos pertinentes
4. Si el lugar de destino no es un Incendio ENTONCES
 - 4.1 Activar Recurso (ver operación propia)
 - 4.2 Actualizar el valor de la matriz ráster en ese lugar (incrementarlo)
5. Si el lugar corresponde a un INCENDIO
 - 5.1 Asignar el recurso al incendio (ver operación propia)
6. Actualizar listas de recursos (ver operación propia)
7. Restaurar el modo de operación a INFORMACION
8. Repintar el mapa

Tabla 16: Colocar/Recolocar un recurso en el mapa.

Activación de un recurso inactivo	Interna
Se trabaja con la lista general de recursos (el objeto recurso que en concreto se activa) y con las listas de índices que deberán ser actualizadas después de realizar esta operación	
Internamente cuando se coloca un recurso en una posición que no corresponde a un incendio (sigue siendo un recurso activo libre).	
Cambia los datos de posición y estado del recurso afectado	
<ol style="list-style-type: none"> 1. Recuperar el objeto recurso de la lista general de recursos <ol style="list-style-type: none"> 1.1 Recuperar su índice de los índices de inactivos 1.2 Recuperar el objeto de la lista a partir de su índice 2. Hacer las modificaciones previstas <ol style="list-style-type: none"> 2.1 Poner el estado a ESTADOACTIVO 2.2 Actualizar su posición <ol style="list-style-type: none"> 2.2.1 Colocarlo en el nuevo municipio destino 2.2.2 Colocarlo en la nueva posición dentro del municipio 	
NOTA: posteriormente al actualizar listas quedará reflejada esta activación en la lista de índices	

Tabla 17: Activación de un recurso inactivo.

Desactivación de un recurso (asignado o no)	Interna
Principalmente la lista de recursos de la Comarca y las listas de Indices. Si el recurso está asignado a un incendio también se ve involucrado el objeto correspondiente a él.	
Internamente se realiza cuando se recoloca un recurso activo. El usuario puede desencadenar esta acción cambiando el estado del diálogo del recurso a Fuera de Servicio	

Se distingue entre la desactivación de un recurso libre activo o un recurso asignado a un incendio. En el primer caso se accede al recurso a partir de su índice y se invalidan los datos de estado y posición, a su vez, actualizamos la posición de la matriz *ráster* del municipio dónde estuviese (puede no ser el municipio actual). Si está asignado además de todas las operaciones ya mencionadas se debe modificar la lista de índices de recursos asignados del incendio en cuestión

CASO 1: El Recurso no está asignado a ningún incendio

1. Recuperar los datos del recurso de la lista general de la comarca
 - 1.1 Obtener el índice del recurso de la lista de índices activos
 - 1.2 Obtener el objeto del recurso
2. Cambiar el estado a INACTIVO
3. Destactivar su colocación en el mapa
 - 3.1 Se coloca su posición en INVALIDA (municipio y Posición)
 - 3.2 Se desactiva la posición *ráster* del municipio donde estaba (sólo si no estaba asignado a un incendio)
4. Reorganizar las listas de índices (Activos e Inactivos).

CASO 2: El Recurso está asignado a un incendio.

1. Recuperar el incendio de la lista de la comarca
2. Desactivar Recurso (CASO 1)
3. Actualizar las lista de índices de recursos asignados al incendio

Tabla 18: Desactivación de un recurso.

Consulta/Modificación de información de un recurso	Entrada/Salida
	La clase Fuego y la Clase ListaFuegos. Afecta al objeto ListaFuegos del Objeto Comarca eliminando el Fuego correspondiente. También afecta a la lista de Recursos de la Comarca cambiando a estado ACTIVO los que estaban en estado ASIGNADO y en ese incendio.
	Se activa haciendo doble <i>click</i> sobre el recurso en cuestión en la lista de recursos donde aparezca la lista de activos, de Inactivos, y de Asignados a un incendio.
	Devuelve todos los datos del diálogo a la Base de Datos y marca estos datos como finalizados (no de trabajo inmediato), además elimina toda presencia del incendio en la EI (lista de incendio y matriz <i>ráster</i>) y deja allí los recursos que ha tenido asignados para que sean luego recolocados donde proceda.

1. Si no hay Recurso Seleccionado entonces TERMINAR
2. Recuperar los datos del recurso de la lista general de la comarca
 - 1.1 Obtener el índice del recurso de la lista de índices correspondiente a la lista que se haya utilizado(activos, inactivos y asignados).
 - 1.2 Obtener el objeto del recurso
2. Según el tipo de recurso se presenta un diálogo de información diferente
 - 2.1 Inicializar los datos del diálogo (identificador del recurso)
 - 2.2 Cargar el diálogo
 - 2.3 OPERACIONES DEL USUARIO
4. Volcado de datos del diálogo a la BD
3. Si el recurso ha pasado a estar "FUERA DE SERVICIO" entonces Desactivar Recurso (ver operación propia)
4. Actualizar las listas de recursos correspondientes al lugar
5. Si no se está en el diálogo de un incendio entonces Repintar Mapa

Tabla 19: Consulta/Modificación de información de un recurso.

Actualización de las listas de recursos	Interna
Se consultan la lista de Recursos de la Comarca y las listas de índices para distinguir entre Activos e Inactivos	
Se activa cada vez que se ha producido un cambio en los recursos procedente de alguna operación del usuario	
Revisa el estado de todos los elementos de la lista y según su estado va insertando sus índices en el vector correspondiente y el ícono del tipo y su nombre en la lista que le corresponda. Necesaria para mantener la coherencia de las listas de imágenes	
<ol style="list-style-type: none"> 1. Resetear los controles Lista <ol style="list-style-type: none"> 1.1 Resetear a lista de recursos Inactivos <ol style="list-style-type: none"> 1.1.1 Borrar todos los ítems 1.1.1 Crear de nuevo la lista de imágenes relacionadas 1.2 Resetear la lista de recursos Activos <ol style="list-style-type: none"> 2.1.1 Borrar todos los ítems 2.1.1 Crear de nuevo la lista de imágenes relacionadas 1.3 Para cada recurso de la lista general de la comarca <ol style="list-style-type: none"> 1.3.1 Si su estado es inactivo <p>entonces Insertar su índice en los índices de Inactivos sino Insertar su índice en los índices de Activos</p> 1.3.2 Según el tipo del recurso <p>Insertar su ícono en la lista de imágenes correspondiente</p> 1.3.3 Según su estado <p>Insertar su estado en el control lista correspondiente</p> 	

Tabla 20: Actualización de listas de recurso.

Repintado del mapa	Interna
	Se consultan los datos presentes en la matriz ráster y algunos datos temáticos como recurso activo seleccionado, estado del incendio...
	En diversas operaciones donde las modificaciones afectan a la matriz <i>ráster</i> siempre se invoca esta operación.
	Coloca gráficamente los elementos georreferenciables en su lugar del mapa, el número de recursos en cada unidad cartografiable, dónde está el recurso actualmente seleccionado, y los incendios con un ícono coloreado según el estado de éste.
1. Preparar el dispositivo de dibujo	
1.1 Operaciones MFC	
1.1.1 CPaintDC dc(this);	
1.1.2 CWnd *pWnd = GetDlgItem(IDC_MAPA);	
1.1.3 CDC* pControlDC = pWnd->GetDC();	
1.1.4 pWnd->Invalidate();	
1.1.5 pWnd->UpdateWindow();	
1.2 Establecer los parámetros para pintar el mapa	
1.2.1 Construir el punto inicial	
1.2.2 Construir el objeto indicador el tamaño	
2. Pintar el mapa cartográfico base del municipio activo	
3. Representar en el mapa los valores de la matriz ráster	
Para todos los puntos de la matriz	
3.1 Definir el rectángulo de dibujo	
3.2 Segundo el valor de la matriz	
PRESENCIA/AUSENCIA: No pintar nada	
INCENDIO: Existe un incendio en esa posición	
3.2.1a Segundo estado del incendio elegir ícono	
3.2.2a Dibujar el ícono en el mapa	
RECURSO: Hay recursos en esa posición	
3.2.1b Representar el número de íconos en la posición	
4. Si hay recurso Activo seleccionado	
4.1 Destacar con un punto amarillo su posición en el mapa	

Tabla 21: Repintado del mapa.

3.8. Descripción funcional del Sistema.

Como se ha señalado en los apartados anteriores, con el acelerado desarrollo de la computación, la percepción remota y la cartografía digital en los últimos decenios, surge la tecnología de los Sistemas de Información Geográfica, que se ocupan de la adquisición, análisis y representación de información georeferenciada. En la actualidad, estos sistemas juegan un papel primordial para la toma de decisiones en las más diversas esferas de la vida del hombre; no obstante, la carencia de herramientas de modelación-simulación, Sistemas Expertos y técnicas de decisión

integradas en los SIG, limitan la potencialidad de los mismos para cumplir con mayor efectividad su misión. INCEND-IA es una propuesta de Sistema Espacial con Soporte a la Decisión como vía idónea para integrar tales tecnologías, y se construye un SIG con técnicas de evaluación multicriterio, decisión multiobjetivo y razonamiento con incertidumbre mediante Lógica Borrosa.

En el modelo de ayuda a la decisión del sistema INCEND-IA se presta especial atención a la previsión y asignación de medios en los incendios forestales, se presenta el problema de cómo reorganizar los medios disponibles en una zona cuando parte de ellos están ocupados en incendios, con el objetivo de dejar esa zona óptimamente cubierta. La solución que se aplica consiste en definir puntos relevantes en el mapa de la zona y mediante interrelaciones entre los puntos a cubrir y los objetos disponibles, obtener una situación equilibrio óptimo, sin perder de vista las necesidades y heurísticas de los propios expertos.

4. Mapa de la aplicación.

INCEND-IA es un sistema integral para la gestión y predicción de incendios forestales que contiene elementos de los Sistemas de Información Geográfica y un fuerte componente de Sistema de Ayuda a la Decisión.

Con el fin de facilitar al usuario el manejo del sistema se ha desarrollado para ser utilizado en entornos Microsoft Windows abarcando las versiones Windows 9x y Windows NT. Por ello se ha pretendido que el interfaz sea lo más gráfico y amistoso posible además de permitir que la aplicación posea la apariencia estándar de todas las aplicaciones de Windows, por lo que posee las barras de menús y estado propias de este entorno. A partir de la ventana principal que aparece cuando se ejecuta, se pueden seleccionar las diferentes regiones entre las que se va a tratar pudiendo elegir entre el mapa general o una zona determinada accediendo al menú *Municipio* (fig.10).

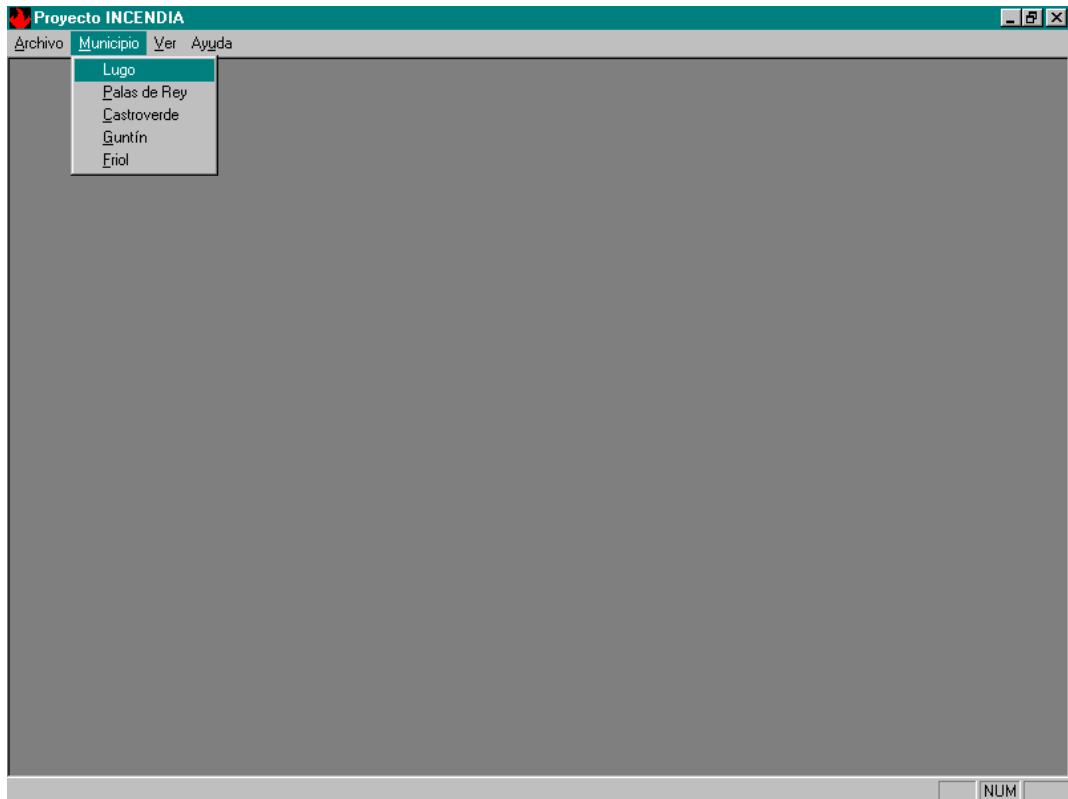


Figura 10: Menú Municipio.

Las ventanas que presentan los municipios constan del mapa del municipio correspondiente dividido en cuadrículas que se corresponden con un Km² de superficie real y una representación analógica de las principales vías de comunicación, núcleos urbanos y masas forestales importantes, así como de las zonas limítrofes. Al lado del mapa aparecen dos listas de Recursos: Una en la parte superior destinada a los recursos inactivos y otra en la parte inferior que contiene la información de los recursos activos. Estas listas permiten al usuario que mediante el sencillo uso del ratón seleccionar un recurso que no está trabajando e insertarlo en el mapa en el punto donde está situado.

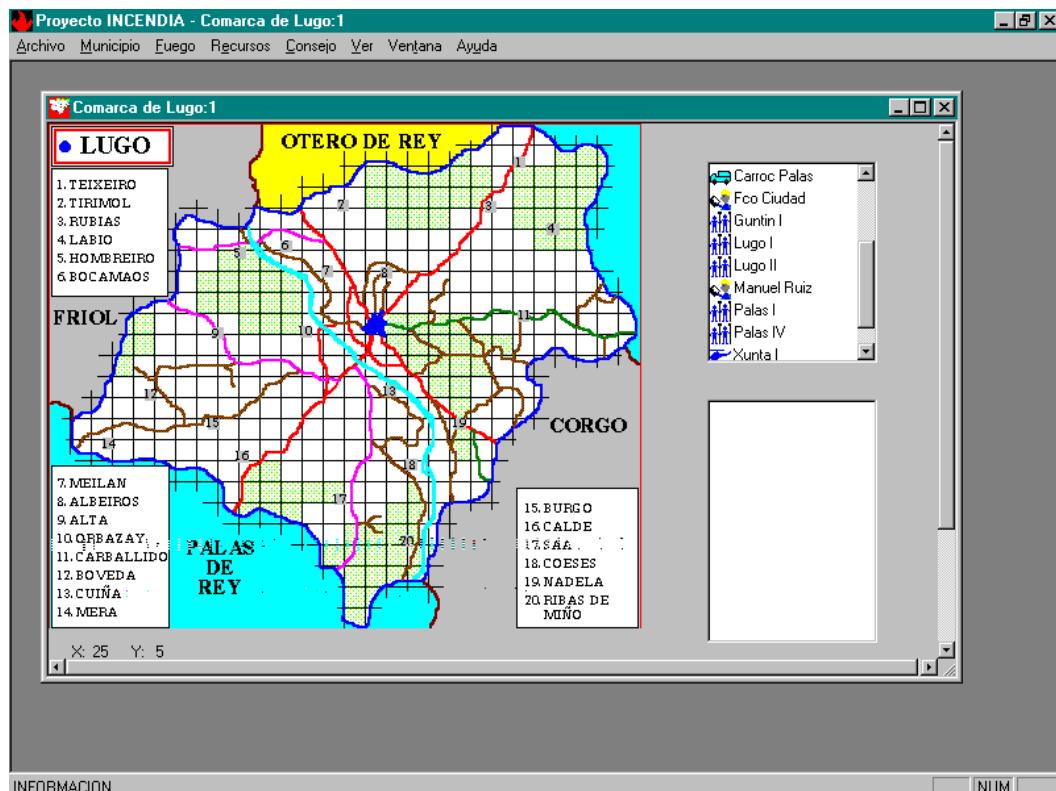


Figura 11: Ventana Municipio.

Como se ha dicho, el sistema INCEND-IA se puede clasificar como una aplicación MDI (*Multiple Document Interfaces*) lo que implica que se pueden estar manejando simultáneamente varios municipios pudiendo visualizarlos simultáneamente en pantalla.

4.1. Operaciones de gestión de Incendios

4.1.1. Inicio de un incendio

La aplicación permite ubicar los incendios en el lugar donde se producen e incluir en ellos los recursos que se encuentran allí para sofocarlos. Para ubicarlos dentro de un municipio se debe recurrir a la opción de menú Fuego+Inicio (fig. 12) para posteriormente hacer *click* con el ratón sobre la posición dentro del municipio donde se haya iniciado el incendio. Si no se hace *click* sobre una posición válida se presentará un mensaje de error. En otro caso se le presentará al usuario un diálogo de introducción de los datos característicos del incendio (fig. 13):

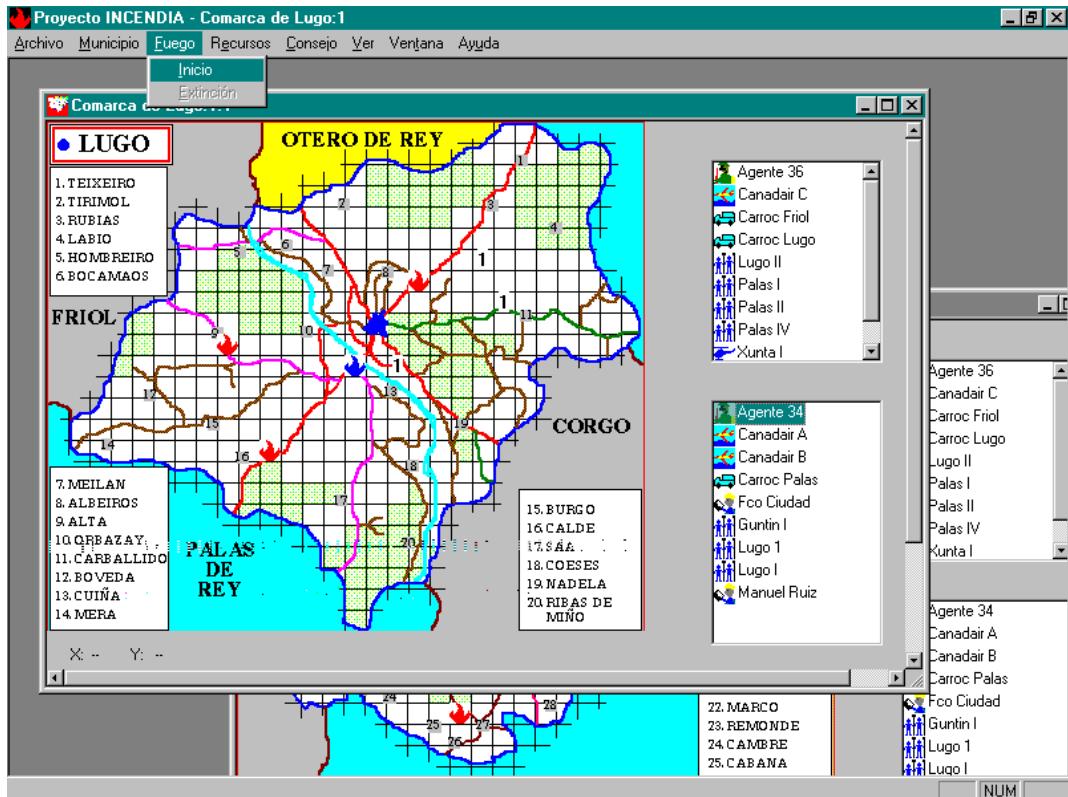


Figura 12: Inicio de un incendio.

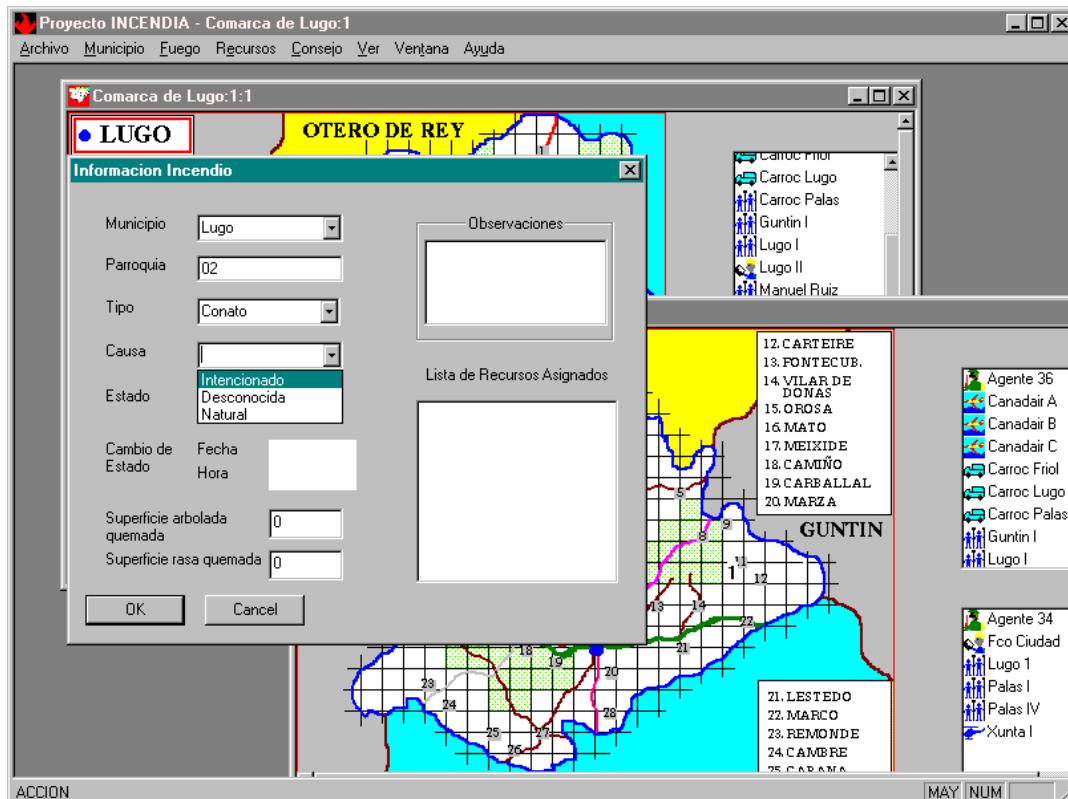


Figura 13: Datos de un incendio.

1. Municipio donde se ha producido.
2. Parroquia donde se ha producido.
3. Tipo (Conato, Quema de Matorral, Incendio).
4. Causa (Desconocida, Intencionada, Natural).
5. Estado.
6. Observaciones.
7. Superficie Arbolada Quemada.
8. Superficie Rasa Quemada.

La lista de recursos asignados es el reflejo de las operaciones de asignación de los recursos al incendio, pudiéndose consultar la información de éstos (ver sección *Operaciones con Recursos*).

Una vez concluida la introducción de datos se debe pulsar OK pudiéndose observar en el mapa la presencia de un ícono representativo del incendio en la posición que se ha indicado (fig. 14).

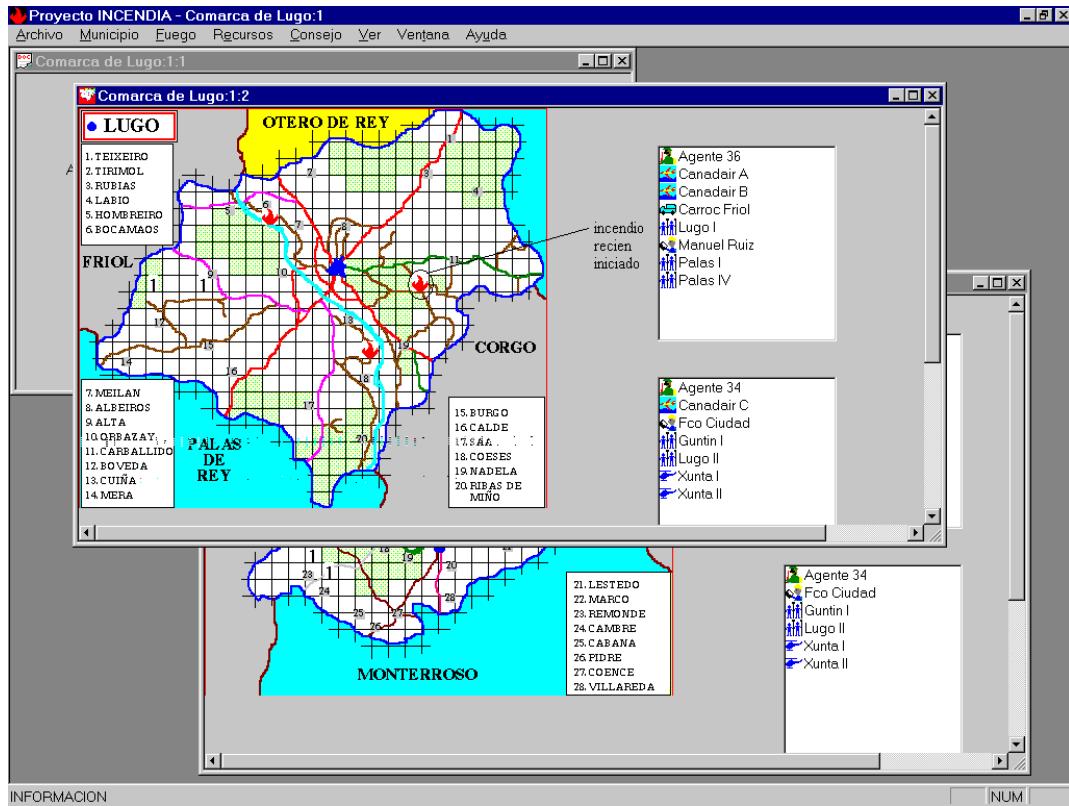


Figura 14: Incendio iniciado.

4.1.2. Consulta/modificación de la información del incendio

Para acceder a la información característica de un incendio se debe activar el diálogo correspondiente haciendo doble *click* con el botón izquierdo del ratón en el ícono representativo del incendio. Una vez abierto el diálogo se puede consultar y modificar toda la información que se ha introducido anteriormente. Para que las modificaciones realizadas tengan efecto se debe pulsar OK para abandonar el diálogo.

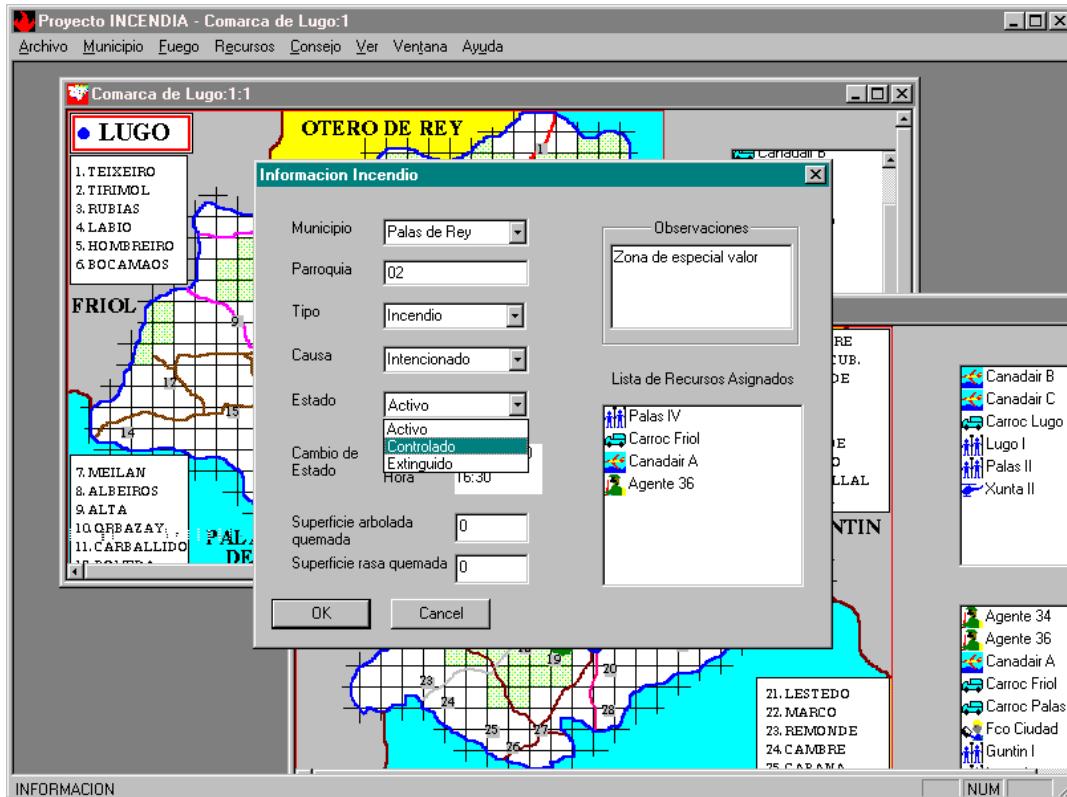


Figura 15: Datos de un incendio: Modificación de estado.

Merece mención aparte el cambio de estado del incendio ya que provoca cambios en la representación de éste dentro del mapa. Por ejemplo, si el incendio ha sido controlado se debe pasar de Activo a Controlado, lo que se realiza seleccionando la opción correspondiente de la lista desplegable de la etiqueta Estado (fig. 15). Una vez realizado este cambio se presenta un diálogo que requiere la introducción de la fecha y la hora de este cambio de estado para que conste en las bases de datos (fig. 16):

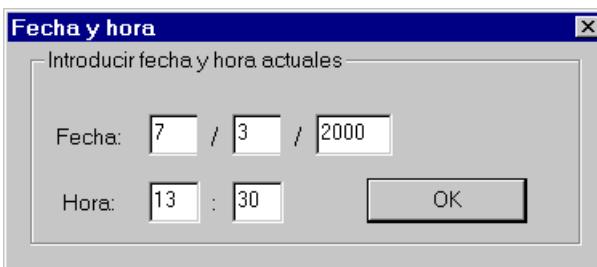


Figura 16: Datos de un incendio: Fecha y hora de la modificación de estado.

Una vez terminada la consulta de los datos del incendio y habiendo modificado su estado se presenta con un color azulado dentro del mapa (fig. 17):

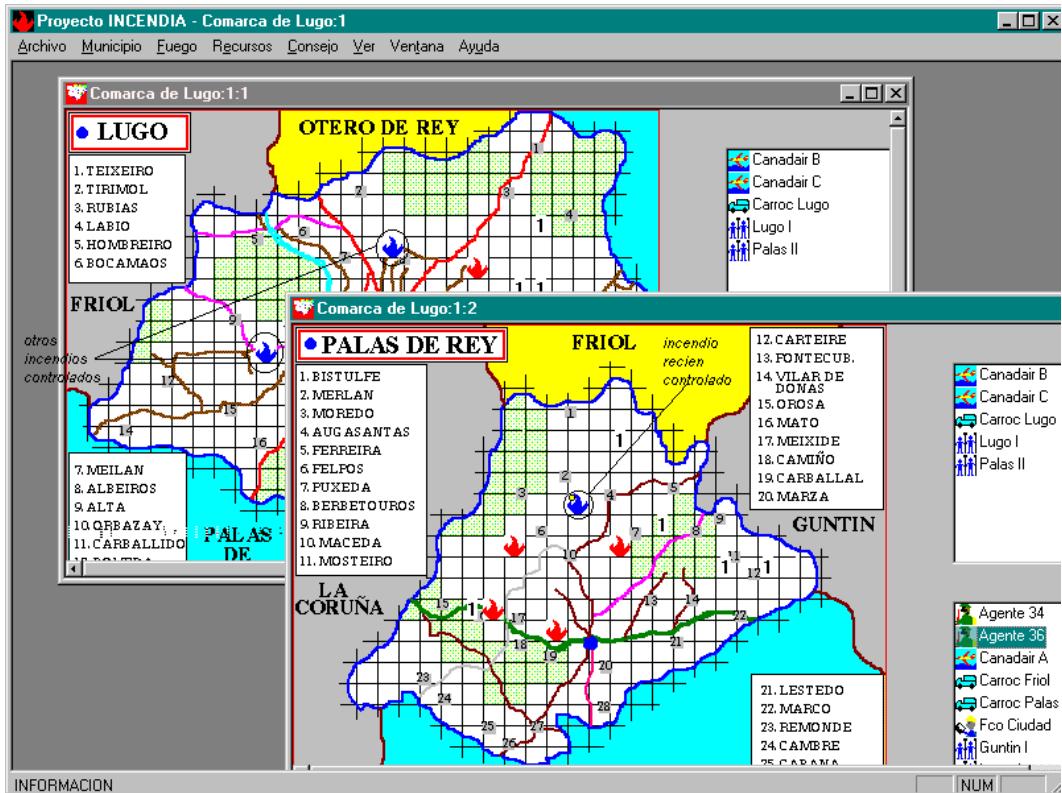


Figura 17: Cambios de color al cambiar de estado un incendio.

Para reflejar la extinción de un incendio se procede de la misma manera que para cambiar su estado (fig. 18):

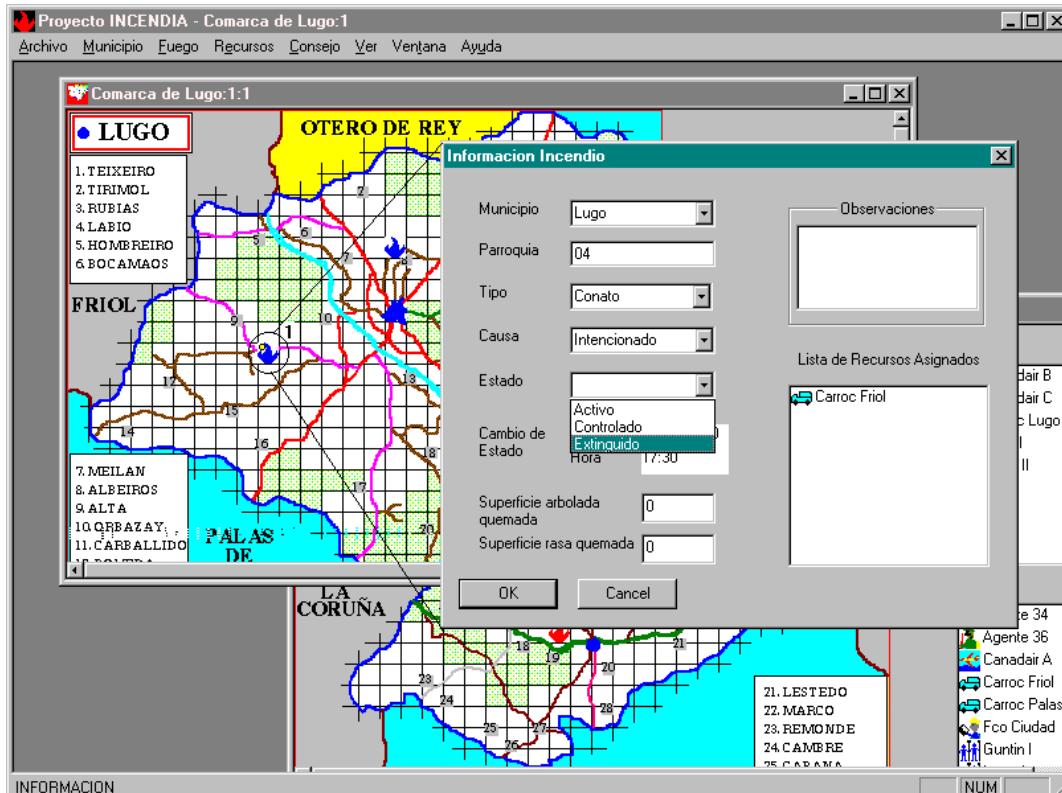


Figura 18: Extinción de un incendio.

1. Se hace doble *click* sobre el incendio en cuestión.
2. Dentro del diálogo de información se cambia su estado a extinguido.
3. Se establece la fecha y la hora de extinción.
4. Se pulsa OK para salir del diálogo.
5. Se observa que el ícono del incendio ha desaparecido quedando en esa posición un indicador de los recursos presentes que están a la espera de ser reasignados.

4.2. Operaciones con los recursos

Los recursos están representados por el conjunto “ícono representativo del tipo + nombre” dentro de las diferentes listas de recursos presentes en el sistema en las que se localizarán según sea su estado: Si está inactivo estará en la lista de recursos inactivos de la ventana del municipio, si esta asignado a un incendio en la lista de recursos asignados del diálogo de información del incendio...

Desde cualquier lugar se puede consultar la información sobre un determinado recurso con tal de pulsar doble *click* con el botón izquierdo del ratón sobre su representación en cualquier de las listas del sistema.

Al consultar la información de un recurso se pueden distinguir varios casos:

1. *Consulta sobre un recurso de la lista de Inactivos:* Se presenta una ficha común para todos los tipos de recursos con información genérica.

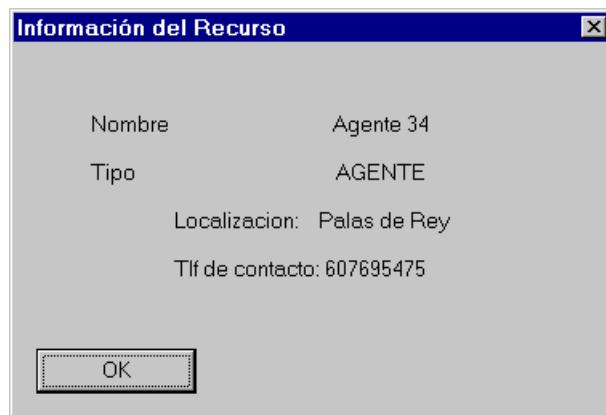


Figura 19: Ficha de recurso inactivo.

2. *Consulta de un recurso de la lista de Activos:* Se presenta un diálogo diferente para los tipos de recursos que se manejan: Medios Humanos (brigadas, agentes y técnicos), Medios Aéreos (helicópteros y aviones Canadair) y Vehículos Terrestres (motobombas). En ellos se puede consultar y modificar mediante distintos cuadros de edición los datos relevantes a la actuación de estos recursos como avituallamiento, repostaje, horarios... (fig. 20).

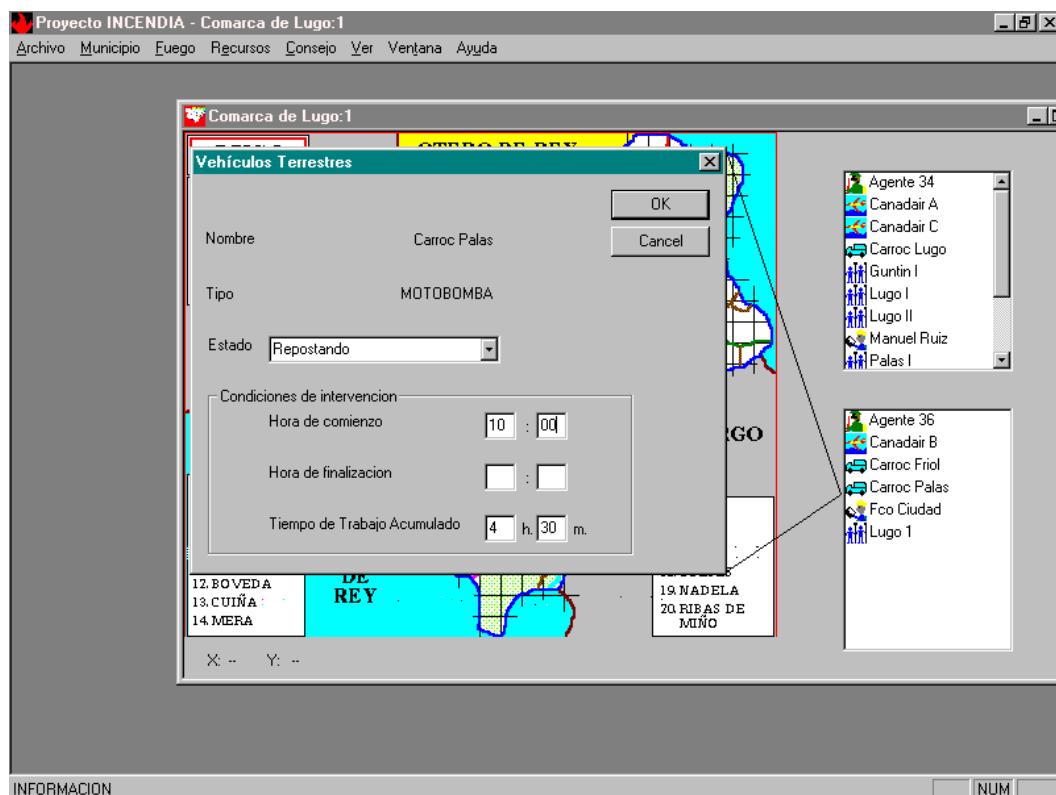


Figura 20: Ficha de recurso activo.

3. Consulta de un recurso de la lista de Asignados dentro del diálogo de Información del incendio: A pesar de realizarse en un lugar distinto las características de la consulta son idénticas al caso anterior (figs. 21 y 22).

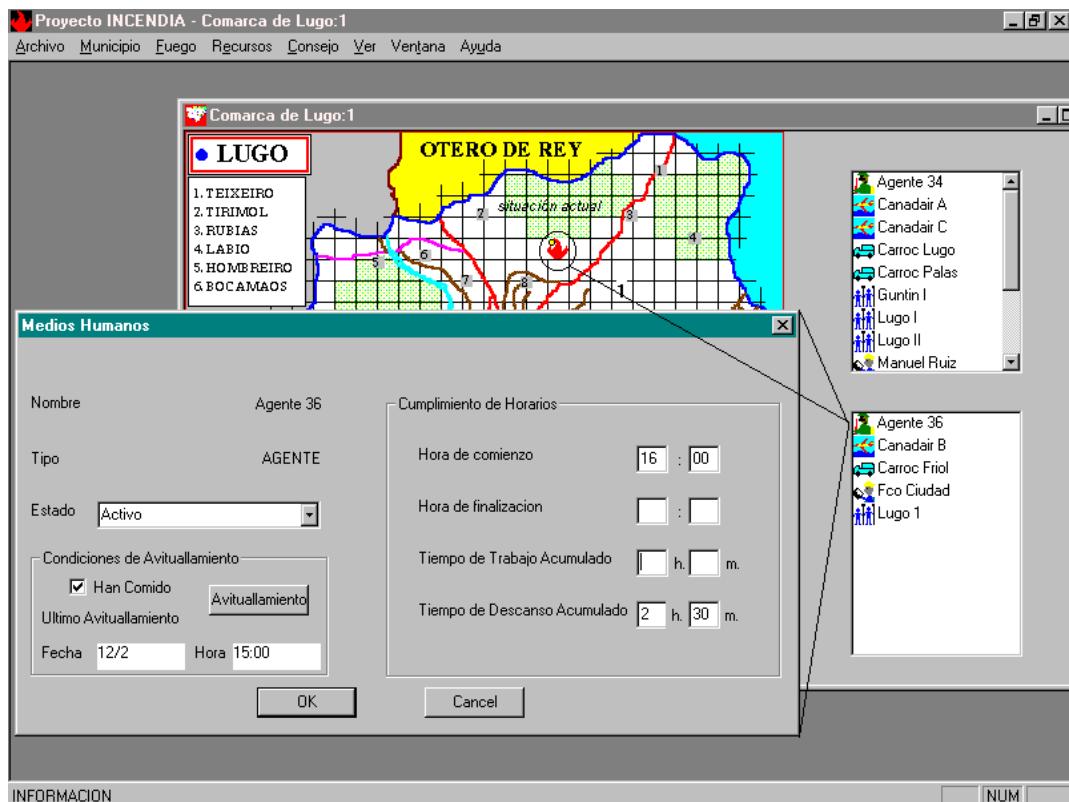


Figura 21: Ficha de recurso activo.

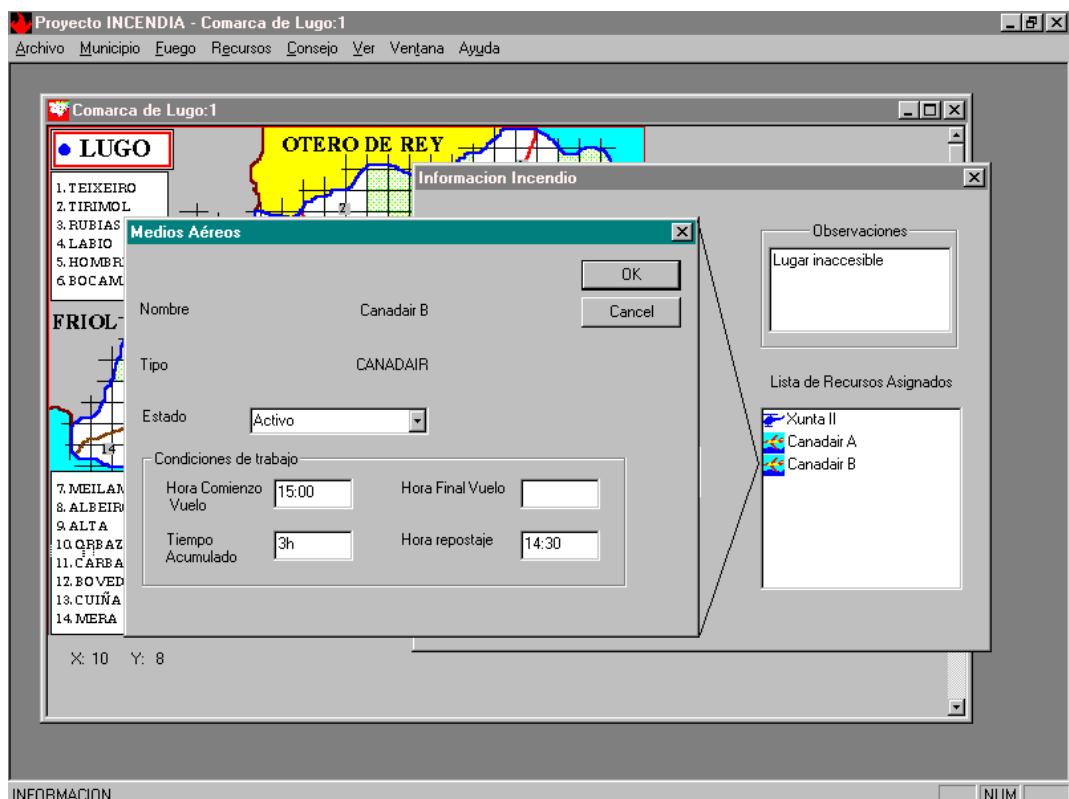


Figura 22: Ficha de recurso activo.

Dentro de las operaciones de cambio de datos que se pueden ejecutar dentro de los diálogos, existe la posibilidad de que el recurso Activo pase a ser un recurso Inactivo. Para ello dentro del diálogo del recurso se debe poner la opción Estado del Recurso a Fuera de Servicio, completar la modificación de los datos y pulsar sobre el botón OK del diálogo.

4.3. Modulo de consejo.

El módulo de consejo (ayuda a la decisión), posee entre otras características, la de predicción de número e intensidad de incendios y el número de recursos a utilizar en una zona determinada. Esta opción se activa desde el menú *Consejo+Predicción*.

Para ello en el primer día después de una época de lluvia se deben introducir los datos de los factores de ciclo, pulsando previamente en el botón Comienzo de Ciclo e introduciendo mediante barras de desplazamiento los distintos valores que correspondan a la situación actual en cuanto a Sequedad estacional, Intensidad de la última Lluvia e Influencia Social.

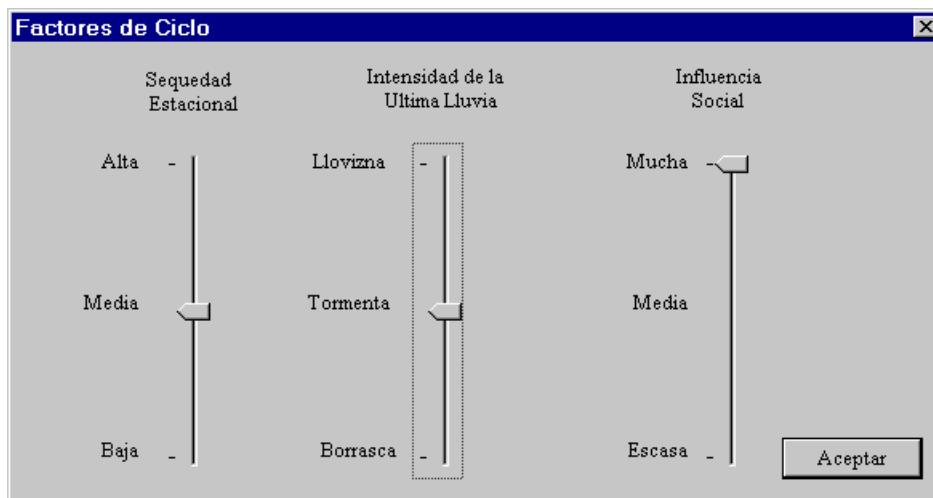


Figura 23: Factores de ciclo.

El resultado será una tabla con los valores previstos para el número de incendios, y el número de recursos que se van a utilizar durante el ciclo.

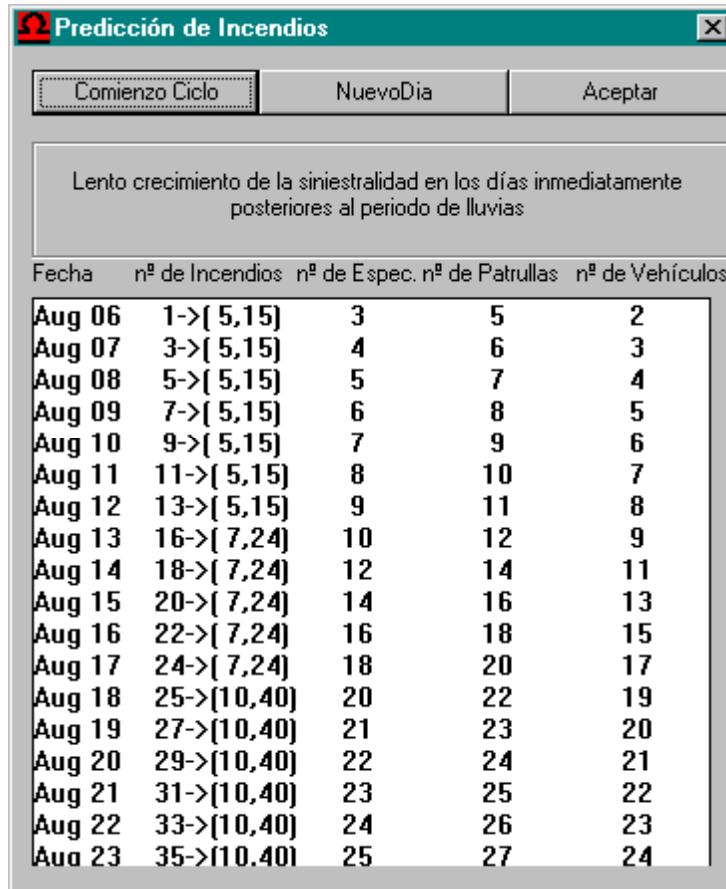


Figura 24: Previsión de los próximos días.

Diariamente se deberán introducir los factores diarios cualitativos haciendo *click* sobre el botón Nuevo Día del interfaz de predicción. El usuario deberá indicar los valores cualitativos de Virulencia de los incendios del día anterior y en cuanto a Sequedad Local (probabilidad de Ignición). También deberá indicar si se presenta alguna situación especial en debido la Previsión de Lluvias o Influencia Social Diaria (fig. 25).

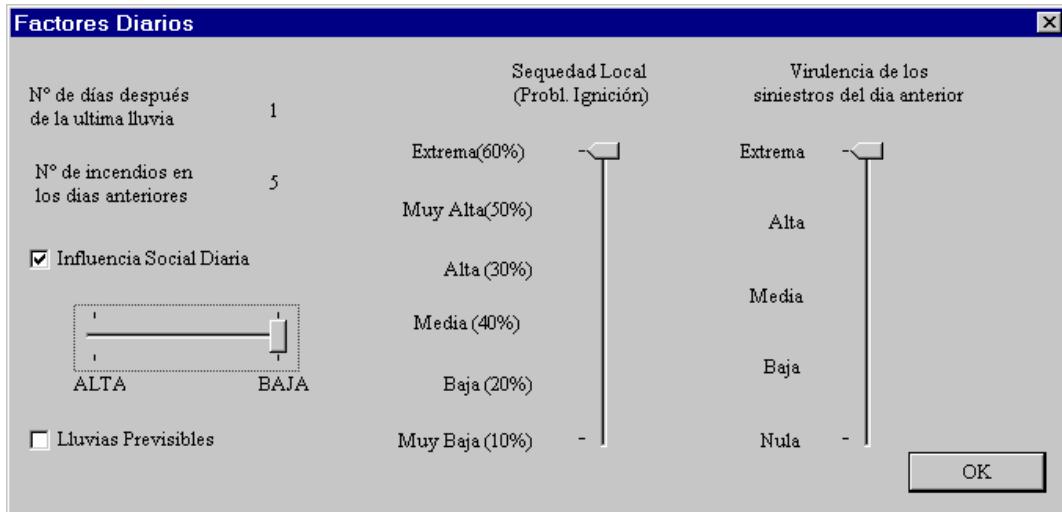


Figura 25: Factores diarios.

El sistema también ayuda a la resolución del problema de reorganizar los recursos disponibles en una zona para lograr que esté cubierta óptimamente. Para ello, se debe activar la opción de menú Consejo+Redistribución de recursos presentándose un interfaz como el que sigue (fig. 26):

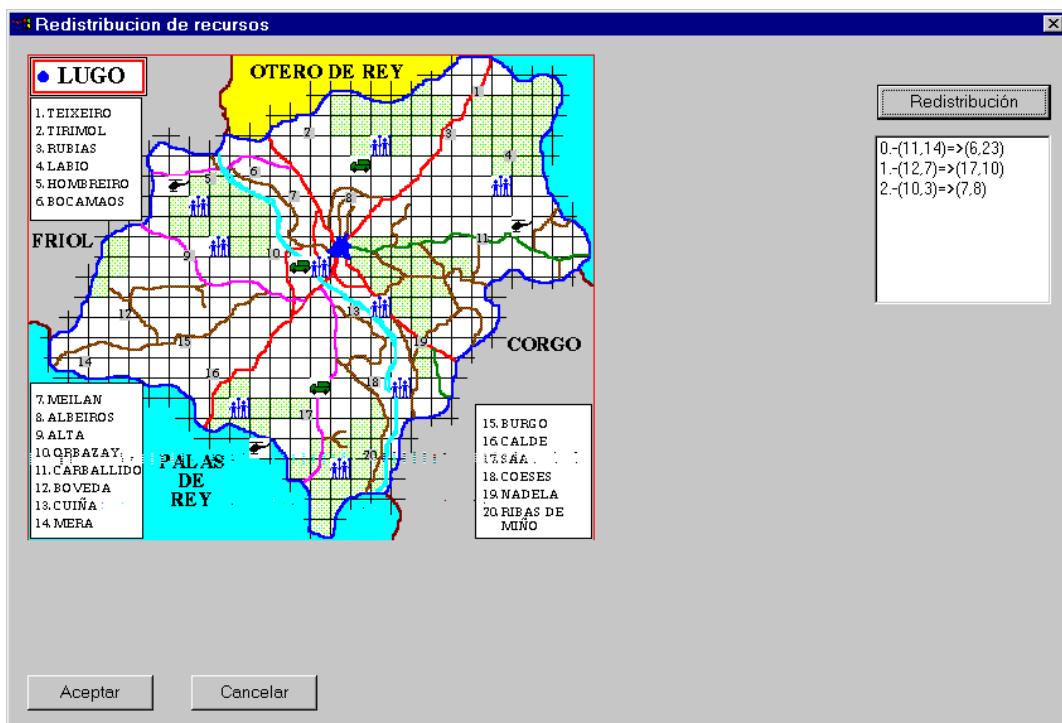


Figura 26: Redistribución de recursos.

Una vez allí y pulsando el botón Redistribuir se podrán observar, tanto en el mapa como en la lista de movimientos, los recursos que se deberán mover para conseguir una cobertura óptima de la zona (fig. 26).

Capítulo 4

Evaluación de Casos Reales.

En este capítulo se muestra la aplicación concreta de extracción de conocimiento prototípico (DCPB) de una base de datos sobre Incendios Forestales, es decir, el Descubrimiento de Conocimiento Prototípico Borroso en datos sobre Incendios Forestales, con los pasos que esto conlleva:

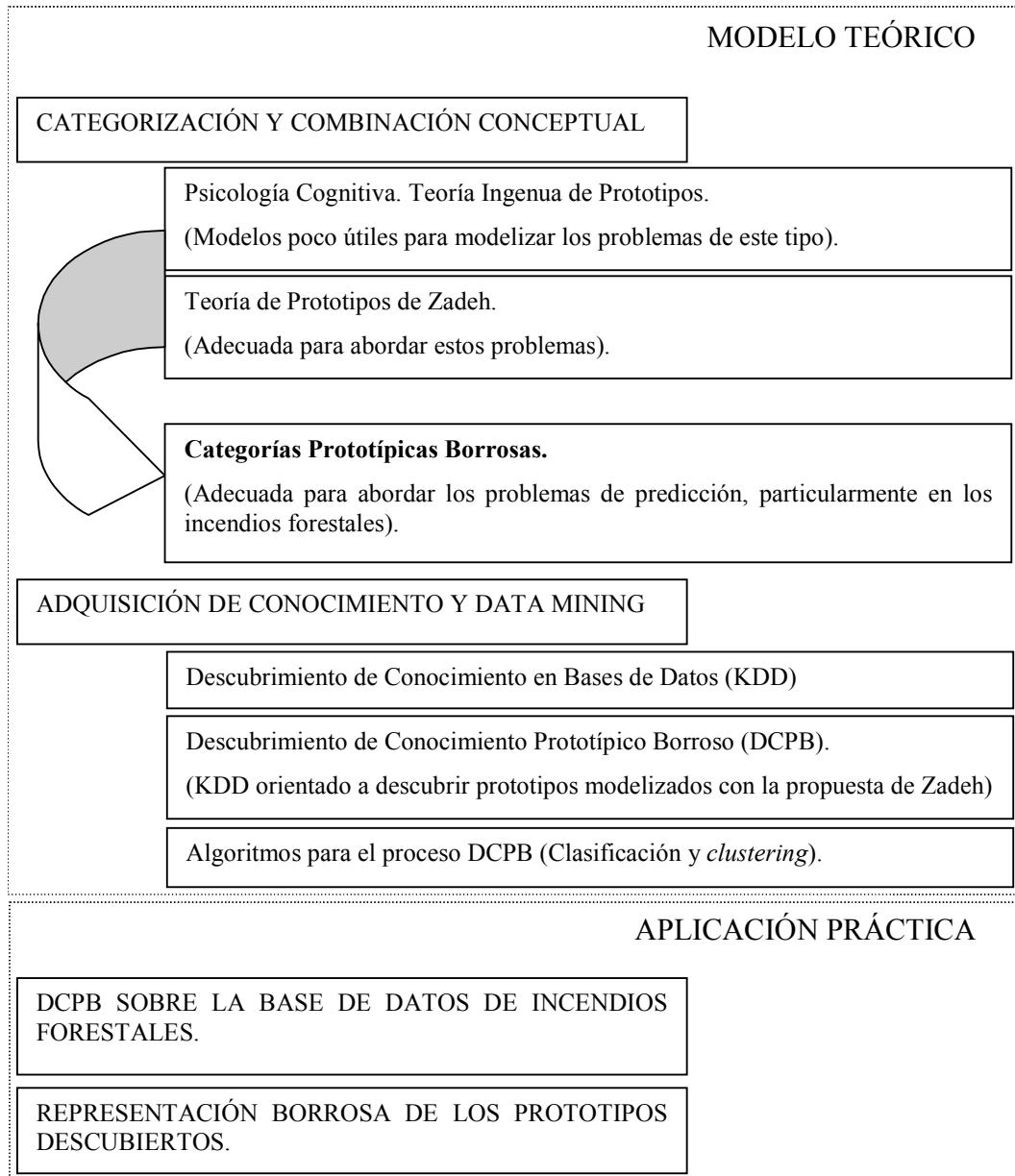
- Selección.
- Preproceso. Eliminación de ruido.
- Transformación.
- *Data Mining*.

A continuación se lleva a cabo la representación borrosa de los prototipos, usando números borrosos triangulares y, bajo otro punto de vista, utilizando los algoritmos de inducción ID3 y CART.

Posteriormente se realiza el *Matching* entre el Caso Real y las Categorías Prototípicas Deformables Borrosas en las siguientes etapas:

- Aproximación (deformación) de las Categorías Prototípicas Deformables Borrosas al ciclo real.
- Aproximación del ciclo real a la situación diaria (Predicción).

Finalmente se presenta un método de redistribución de recursos.



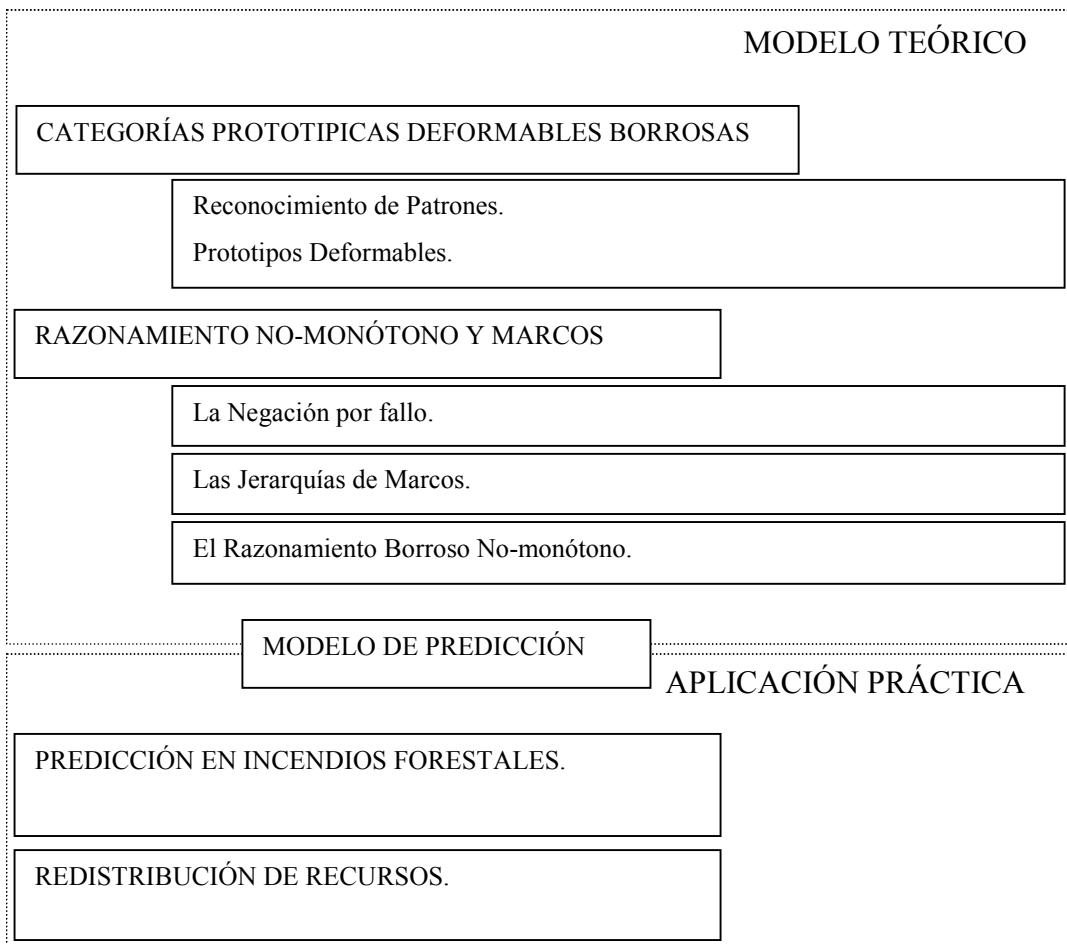


Figura 27: Diagrama de la estructura del capítulo.

Complejidad de los algoritmos de este trabajo.

Los principales algoritmos de este trabajo tienen la característica de ser heurísticos y/o aproximados. El problema que pueden generar es llegar a una complejidad NP: tiempo polinómico no determinista.

Expresado informalmente, un problema *NP* (*Nondeterministic Polynomial*) sería el que el número de aproximaciones aumenta exponencialmente a partir de los datos de entrada, y debe cumplir una serie de propiedades o características: Certificado conciso (Una vez obtenida una solución es sencillo comprobar si lo es), lo resuelven algoritmos no-deterministas y lo resuelven Máquinas de Turing No-deterministas en tiempo polinómico (aumenta exponencialmente a medida que crece el número de variables). Además para ser *NP-Completos* deben ser NP y que todo problema NP sea polinómicamente transformable a él. Ejemplos de este tipo

de problemas son el del *Recubrimiento mínimo*, el de la *Satisfacibilidad de fórmulas lógicas* o el del *Viajante de Comercio* [Brassard, 90].

Para abordar la NP-Completitud se tiene en cuenta el estudio de las propiedades verificables polinómicamente. Para ello se suele tratar de reducirlos a problemas de tipo P (problemas de decisión que se pueden resolver mediante un algoritmo de tiempo polinómico), que en general son eficientes y cumplen que $P \subseteq NP$.

Como se verá en la descripción pormenorizada de cada algoritmo, los algoritmos “problemáticos” de este trabajo han sido abordados de forma que su complejidad pueda ser controlada para que sean eficientes para los fines propuestos.

1. Adquisición de Conocimiento: Obtención de datos fiables.

1.1. *Incendios forestales en Galicia: Adquisición de Conocimiento y datos.*

El proceso de Adquisición del Conocimiento, es fundamental para la validez real de los métodos que pretenden servir para la resolución de problemas de ayuda en la toma de decisiones de este tipo. Aunque se describen con precisión estos procesos a lo largo de los diferentes epígrafes de este trabajo, parece adecuado abordar previamente ciertos aspectos.

1. Normalmente, la aparición de incendios forestales en la comunidad gallega está ligada a factores socio-económicos, difíciles de cuantificar de una forma precisa, que no se reflejan en los datos estadísticos, ni en los denominados “informes” o “partes”.
2. La medida de factores físicos o meteorológicos no es lo precisa que podría ser deseable. Hay demasiados microclimas, especies arbóreas en poca superficie, pequeños cultivos, ..., como para que características como “humedad relativa” o “estrés hídrico” puedan ser generalizadas a grandes zonas de una forma fiable.
3. En los datos estadísticos disponibles, existe un sesgo fundamental: no reflejan la auténtica peligrosidad e importancia de los incendios ocurridos.

Esto es debido a lo siguiente: Supóngase la aparición de dos alarmas simultáneas en una misma comarca, una de ellas en una zona de matorrales, que habitualmente se utiliza para pastoreo y en la que periódicamente se realizan quemas controladas, sin riesgo por proximidad a una zona de especial valor ecológico, y la otra alarma surge a pocos metros de un bosque con difícil accesibilidad.

Es razonable y habitual, en el caso de que los recursos sean limitados, destinar más medios al segundo caso que al primero, lo que provoca en muchas ocasiones que en las estadísticas posteriores se refleje como importante el primer incendio (ej. 200 Ha. de matorral), y el segundo incluso ni aparezca (el umbral de aparición en estadísticas fue de 0,5 Ha. y posteriormente de 0,1 Ha.), porque se ha atajado en sus inicios.

4. Los criterios de prioridad, por ejemplo para el envío de medios aéreos, no siempre están claros, porque en los puntos donde se toman estas decisiones, la información de la que se dispone no es lo suficientemente objetiva ni completa.

Teniendo en cuenta estos criterios, se ha realizado un análisis previo y “superficial” de todos los “ciclos” de incendios de una comarca durante varios años (Ejemplos en fig. 28) y se ha observado que la evolución de la siniestralidad puede ser representada como una función creciente de tipo sigmoidal dividida en tres sectores y que comenzaría en el día posterior a un periodo de lluvia en el que el número de los incendios se ha reducido a cero.

Este patrón de crecimiento se repite de forma cíclica después de cada periodo de este tipo, pero puede sufrir modificaciones debido a factores específicos.

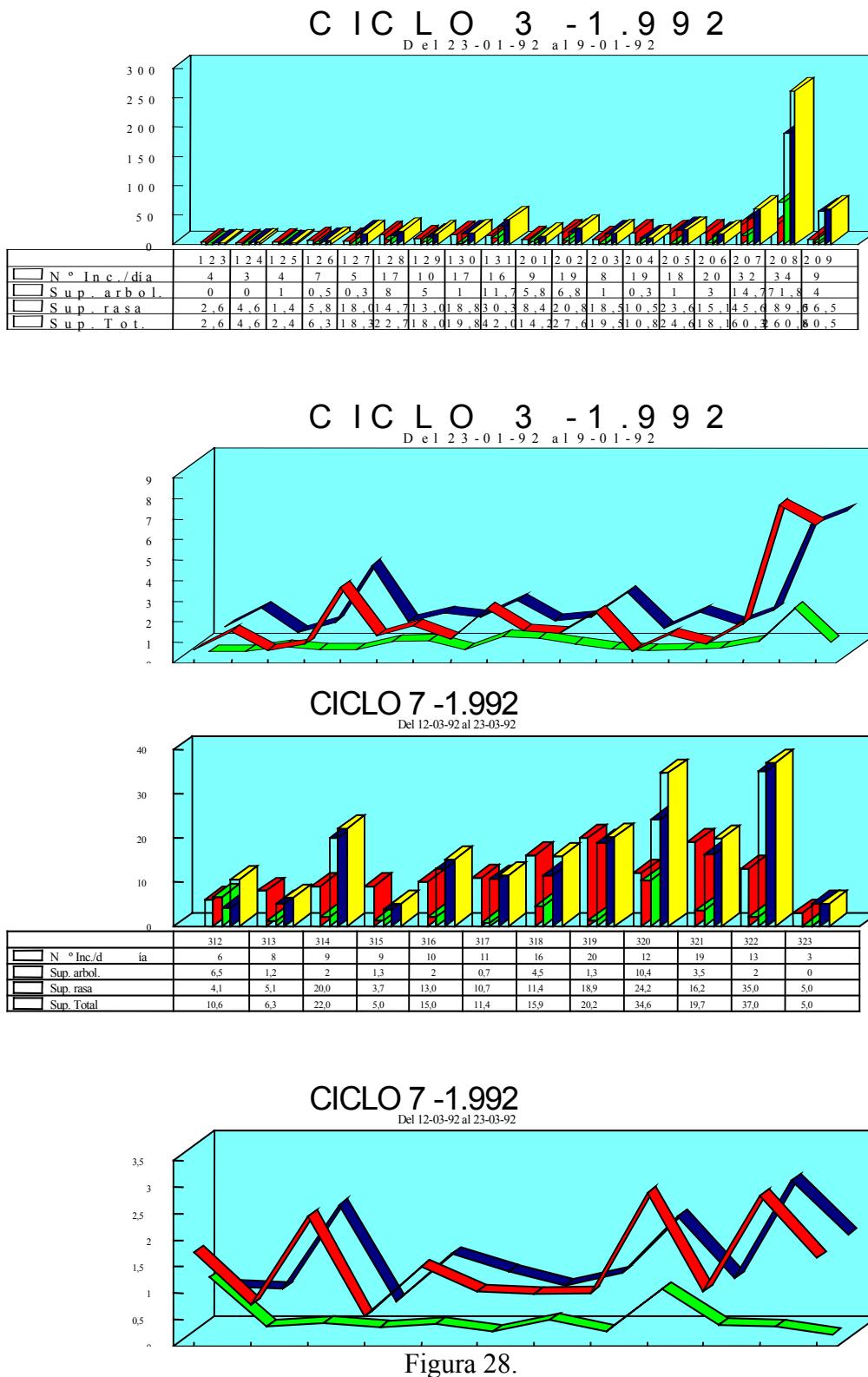


Figura 28.

Entiéndase por siniestralidad la combinación de varios factores, como número, peligrosidad de los incendios, etc. La fig. 29 muestra la representación del patrón de crecimiento.

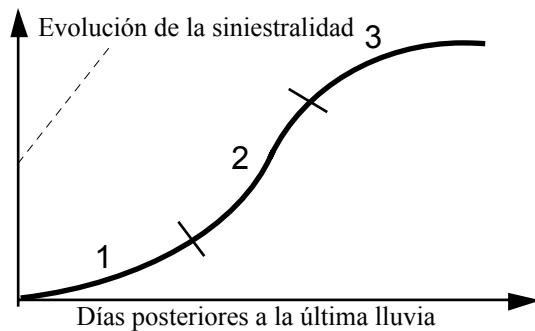


Figura 29.

- El *primer sector* representa un lento crecimiento de la siniestralidad en los días inmediatamente posteriores al periodo de lluvias.
- El *segundo sector* expresa un alto crecimiento de la siniestralidad, especialmente en cuanto al número de incendios.
- En el *tercero* se quiere representar una estabilización en cuanto al número, pero un progresivo aumento en la peligrosidad de los fuegos.

En nuestra vida cotidiana es usual asociar un hecho o conjunto de hechos con un paradigma, de tal forma que el paradigma interpreta la situación y de él dependen las acciones que llevemos a cabo. Por ejemplo, si estamos conduciendo y empieza a granizar adecuamos nuestra conducción al esquema de conducción bajo condiciones potencialmente peligrosas.

Como se ha visto en el ejemplo, muchas de las acciones que realizamos en nuestra vida dependen de una interpretación. Lo que aquí se plantea es que interpretar una situación es encontrar los paradigmas o prototipos afines a las circunstancias del problema. En este caso se trata de simular la capacidad del experto para interpretar la situación, es decir, para encontrar el modelo de evolución de la siniestralidad más adaptado a las circunstancias reales.

1.2. Concepto y Prototipos.

Aunque este tema se ha abordado con profundidad en el capítulo 2, permítase una pequeña nota.

Tomando como marco de referencia la teoría de prototipos de la psicología, podría entenderse que esta representación es prototípica del avance de la siniestralidad.

Sin embargo, en el proceso de adquisición del conocimiento se pudo observar que esta representación simplifica en exceso las pautas del comportamiento de los expertos. Cuando un técnico se enfrenta a una situación real maneja un abanico de prototipos determinados por una serie de factores, es decir, debe decidir qué tipo de evolución de la siniestralidad es previsible. Dicho de otro modo, el prototipo “*Evolución de la siniestralidad*” no es único, sino que existen diferentes formas de evolución dentro de la misma estructura sigmoidal.

Zadeh alude a las teorías clásicas de prototipo desde el punto de vista de la psicología [Zadeh, 1982], criticando precisamente lo que aquí se ha expuesto: su falta de adecuación a la función que debe cumplir un prototipo. La aproximación de Zadeh a lo que debe entenderse por prototipo es menos intuitiva que las concepciones de las teorías psicológicas, pero más racional, más próxima a lo que en un examen detenido muestra el significado de un concepto prototípico.

En nuestro caso se ha observado que la idea de Zadeh sugiere que un concepto engloba un conjunto de prototipos, los cuales representan la buena, baja o media compatibilidad de los ejemplares con el concepto. Las Categorías Prototípicas Borrosas representan las diferentes clases que se pueden determinar en el dominio.

Desde este punto de vista se puede hablar de “*siniestralidad altamente progresiva*”, “*medianamente progresiva*” o “*escasamente progresiva*”. Esto se puede representar, tal y como refleja la fig. 30, con tres sigmoides, entendidas simplemente como representaciones gráficas de los tres prototipos borrosos.

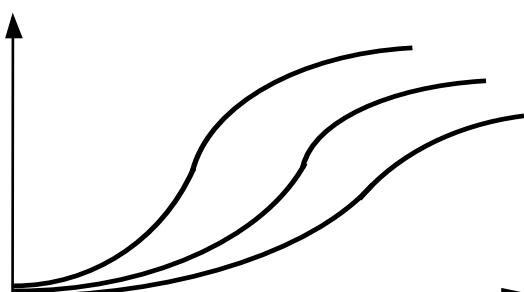


Figura 30.

1.3. DCPB: Descubrimiento de Conocimiento Prototípico Borroso en datos sobre Incendios Forestales.

El proceso de Descubrimiento Prototípico para este caso particular de los datos disponibles sobre incendios forestales, se lleva a cabo en las siguientes fases (fig. 31):

1. *Selección:* Aplicando el conocimiento del dominio y el conocimiento relevante *a priori*, teniendo en cuenta los objetivos del proceso global de DCPB se crea una tarjeta de datos que incluye conjuntos seleccionados de datos o subconjuntos de variables relevantes o ejemplos. Se toma como conjunto de partida una base de datos relacional que contiene aproximadamente 12.000 incendios ocurridos en Galicia durante los años 1991 y 1992. Se seleccionan estos años por ser 1991 un año poco conflictivo y 1992 uno muy conflictivo. Se seleccionan los 3.204 correspondientes a la provincia de Lugo, de los años 1991 y 1992. También se separan por comarcas y se eliminan campos no relevantes. (Como ejemplo, ver Apéndice 4).
2. *Preproceso:* Limpieza de datos, eliminación de ruidos, manejo de campos vacíos, datos perdidos, valores desconocidos o por defecto, evolución de datos. Se aplican técnicas estándar de bases de datos.
3. *Transformación:* Reducción del número de variables: Localización de ciclos de incendios. Búsqueda de formas útiles para expresar los datos dependiendo del uso posterior que se les va a dar y de los objetivos del sistema. Se usa el conocimiento experto y técnicas de transformación e informes en bases de datos. Ordenación y clasificación de los ciclos según la evolución de la siniestralidad.
4. *Data Mining:* Elección de los algoritmos de *Data Mining*. Decisiones acerca del modelo que se deriva del algoritmo de *Data Mining* elegido (clasificación, resumen de datos, predicción). Búsqueda de patrones de interés, en cuanto a clasificación, reglas o árboles, regresión, clasificación, dependencia, heurísticas, incertidumbre. Se generan los prototipos de evolución de la siniestralidad en base a los ciclos representativos de cada una de las clases.

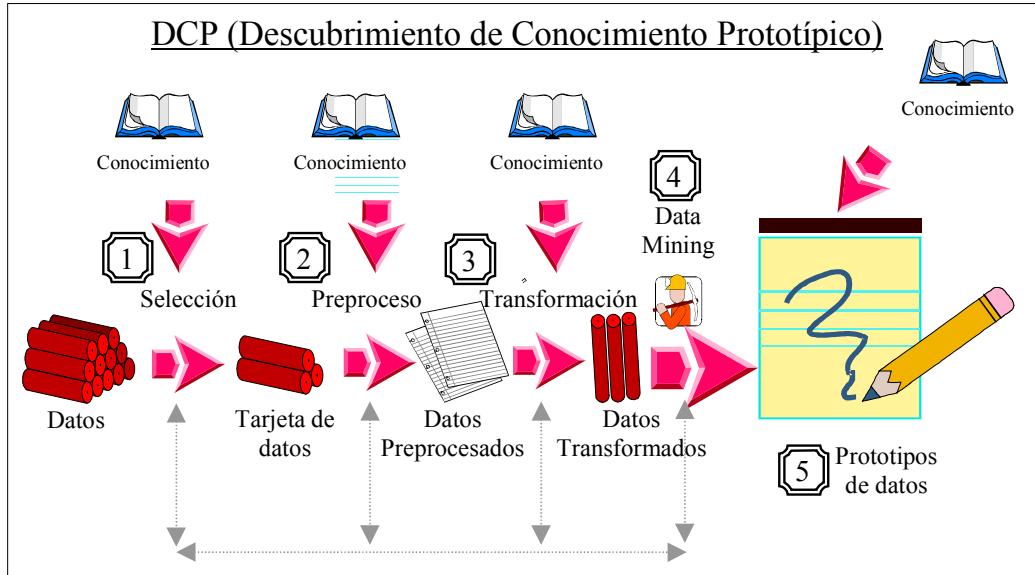


Figura 31: Descubrimiento de Conocimiento Prototípico.

A continuación se muestran con detalle las operaciones de Preproceso, Transformación y *Data Mining*:

1.3.1. Preproceso. Eliminación de ruido.

Una vez extraído de la base de Datos de Galicia los 3.204 incendios correspondientes a los años 1991 y 1992, se debe proceder a su estudio completo, antes de lo cual habrá que hacer unas modificaciones tanto en el diseño como en el contenido de la tabla ya que presenta diversas irregularidades que hacen imposible o muy dificultoso el aprovechamiento de estos datos.

a) Tratamiento de las Fechas y Horas.

La base de datos contiene el control de las fechas y las horas en que los Incendios se han iniciado, se han controlado y se han extinguido. Estos datos son muy importantes ya que permiten el control de la evolución temporal de los incendios y la obtención de datos estadísticos sobre cada día, semanas... En la base de datos inicial estos campos estaban tipificados como simples cadenas de caracteres con un cierto formato, lo cual facilitaba en un primer momento su almacenamiento pero dificulta enormemente su posterior tratamiento ya que pierden toda su semántica propia, no pudiéndose efectuar sobre ellos operaciones como comparaciones, restas o distancias.

La solución a este problema es sencilla, siempre y cuando las fechas y horas hayan sido almacenadas con algún formato estándar, lo que en este caso ocurre, de esta forma cambiando el tipo de los campos en el sistema Gestor de Base de Datos utilizado (Microsoft Access '97), se permite que las aplicaciones accedan y operen sobre estos datos con toda la semántica correspondiente a los tipos Fecha y Hora (*CTime* en Visual C++), tabla 22.

<i>Campo afectado</i>	<i>Formato</i>	<i>Descripción</i>
FECHA_INI	<i>DD/MM/AA</i>	Fecha de inicio del incendio
HORA_INI	<i>HH:MM:SS</i>	Hora de inicio del incendio
FECHA_EXT	<i>DD/MM/AA</i>	Fecha de control del incendio
HORA_EXT	<i>HH:MM:SS</i>	Hora de control del incendio
FECHA_FIN	<i>DD/MM/AA</i>	Fecha de la extinción total
HORA_FIN	<i>HH:MM:SS</i>	Hora de la extinción total.

Tabla 22.

b) *Pérdida de datos numéricos.*

Todos los campos o variables estadísticas que se manejan poseen un conjunto de valores perdidos, es decir, el valor o los valores que no se consideran como válidos para la variable con la que se está trabajando, no existiendo por defecto valores ausentes.

Todos los valores numéricos existentes en esta base de datos son indispensables para el posterior tratamiento de ésta, por ello es necesario que estén bien definidos y se pueda procesar de forma correcta estos valores perdidos.

Existen dos tipos de valores perdidos, los valores omitidos por el usuario (*missing value*) códigos que indican que el verdadero valor de una variable es desconocido y que los casos que contengan esos valores deben ser excluidos del análisis y los valores perdidos por el sistema (*system-missing values*) valores asignados por el SGBD correspondiente cuando un valor de los datos resulta indefinido de acuerdo con el tipo de formato que se ha especificado (como por ejemplo un valor identificativo de *menos infinito*).

En la tabla de incendios que se maneja y debido en su mayor parte a su procedencia, sucede que los valores numéricos han sido introducidos a mano y sin tener en cuenta que debían de tener un valor por defecto, por lo tanto en la

primera parte de la base de datos ocurre que todos los campos están totalmente ocupados por valores que suelen tomar el valor 0 (superficie quemada despreciable, categorías de personal que no estuvieron en el incendio...), el problema surge cuando alrededor de la mitad de la tabla esos valores no aparecen y campos numéricos que deberían tener algún valor no lo tienen, lo cual provoca un cierto caos en su tratamiento ya que el SGBD utilizado y el controlador de ODBC para comunicarlo con la aplicación coloca en estos valores el de *menos infinito* ($-9,18 * 10E-19$) lo cual hace imposible un cálculo automático sobre estos valores si no se controla esta contrariedad (tabla 23).

Campo afectado	(Min,Max)	Tipo de Dato	Descripción
SUP_ARBO	(0,30)	real doble	Superficie arbolada arrasada por el incendio
SUP_RASA	(0,60)	real doble	Superficie rasa afectada por el incendio
SUP_TOTAL	(0,90)	real doble	Superficie total quemada por el incendio
TECNICOS	(0,10)	entero	Nº de técnicos que intervinieron en el incendio
AGEN_FOR	(0,10)	entero	Nº de agentes forestales implicados
AUT_CIVIL	(0,15)	entero	Autoridad civil destinada al incendio

Tabla 23.

Debido a la experiencia acumulada en la gestión de la información que generan los incendios se sabe que estos valores son nulos, despreciables o desconocidos por su poca relevancia, debido a esto se ha procedido a que todos estos casos pasen a tener un valor por defecto 0. Para hacer la actualización de estos campos se ha recurrido al siguiente algoritmo que actúa sobre la tabla mediante ODBC:

```
void CCICLOSDlg::OnLimpiar()
{
    CBDIncendios m_pInc; // acceso a la tabla de incendios

    // inicialización de la Base de Datos
    m_pInc.Open();
    double dSupArb, dSupRas;
    /* BUCLE DE RECORRIDO DE LA BASE DE DATOS */
    while (!m_pInc.IsEOF()) {

        /* se limpian los valores de tipo real de doble precisión */
        if (m_pInc.m_SUP_ARBO > 35 || m_pInc.m_SUP_ARBO < 0) dSupArb = 0;
        else dSupArb = m_pInc.m_SUP_ARBO;

        if (m_pInc.m_SUP_RASA > 75 || m_pInc.m_SUP_RASA < 0) dSupRas = 0;
        else dSupRas = m_pInc.m_SUP_RASA;
```

```

if (m_pInc.m_SUP_TOTAL > 75 || m_pInc.m_SUP_TOTAL < 0) dSupRas = 0;
else dSupRas = m_pInc.m_SUP_RASA;
/* se limpian los datos de tipo entero */
int AgenFor = (m_pInc.m_AGEN_FOR < 10)? m_pInc.m_AGEN_FOR:0 ;
int AutCivil = (m_pInc.m_AUT_CIVIL < 10)? m_pInc.m_AUT_CIVIL:0;
int Tecnicos = (m_pInc.m_TECNICOS < 15)? m_pInc.m_TECNICOS :0;

// se inserta el cambio en la tabla
Insertar(&m_pInc,dSupArb,dSupRas, dSupArb+dSupRas, AgenFor, AutCivil,
Tecnicos);
m_pInc.MoveNext();
}
}

```

Algoritmo 1: Eliminación de valores perdidos.

Complejidad del algoritmo de corrección de datos perdidos y valores desconocidos: $O(N, F) = N \cdot F$ siendo N: número de filas de la tabla y F: número de campos que pueden presentar valores perdidos. La complejidad resulta ser lineal ya que $N \gg F$.

c) Malos diseños

El diseño de una base de datos que gestiona la información sobre un cierto asunto (en este caso incendios) debe ser cuidadoso ya que posteriormente se debe acceder a ellos para obtener resultados de todo tipo. Esta es una premisa que se aleja mucho de la realidad de la base de datos con la que aquí se trabaja.

Esta base datos de una única tabla que gestiona todos los datos necesarios está concebida por su fácil construcción y modificación, sin tener en cuenta para nada los conceptos de modelización y normalización de las bases de datos así como las características esenciales de un buen diseño. Este diseño está más destinado a almacenar de alguna forma sencilla los datos sin ningún interés en una posterior utilización de éstos. Estas situaciones son la pesadilla de las personas dedicadas a la extracción de datos, ya que no se trata de valores perdidos, si no de valores que existen pero no se sabe donde se localizan.

Esta situación se presenta en nuestro caso cuando se quiere reflejar los recursos personales y mecánicos que estuvieron involucrados en la extinción de cierto incendio:

- *Para medios humanos:* 20 campos de tipo texto con los identificadores de cada uno de los recursos. Sólo se ocupan los campos, por orden,

según el número de recursos asignados, el resto de los campos hasta 20 se quedan en blanco. De estos 20 campos, sólo son utilizados con frecuencia los cinco primeros, y como máximo son utilizados 14, por lo que existen campos que sobran.

- *Para medios mecánicos:* De los 20 campos de idénticas características que los anteriores sólo son utilizados alguna vez 10, de los cuales sólo los primeros son utilizados con frecuencia.

El diseño aunque posee un cierto margen para poder albergar muchos recursos asignados a incendios genera espacio desaprovechado, y si de produjera una situación con más de 20 recursos (humanos o mecánicos) la base de datos no podría contemplarlos. Además para obtener datos estadísticos directamente de la base de datos sobre el personal o vehículos utilizados para la obtención del incendio (dato muy relevante) se necesitan hacer comprobaciones que generaría una pérdida de tiempo y quizás también de información que no sería aceptable en un estudio serio de los incendios.

Para no tener que hacer cálculos inútiles e imprecisos sobre estos datos se han creado dos campos nuevos en la tabla: *personal* y *medios*; los cuáles son, respectivamente, un recuento de los medios humanos y mecánicos que han sido asignados al incendio. Estos campos han sido llenados por un algoritmo que ha contado los campos no vacíos de los indicados anteriormente ofreciendo un resultado acumulado de todos ellos. Posteriormente se han eliminado todos los 40 campos que la base de datos destinaba a indicar el personal y los medios.

Los únicos datos perdidos en este proceso son los identificadores correspondientes a los recursos existentes en la comarca, datos que pueden ser obtenidos fácilmente de otras fuentes. Como contrapartida de 40 campos prácticamente inútiles se ha pasado a dos campos numéricos, sin valores perdidos, que ofrecen valores muy significativos para el posterior estudio de estos datos, es decir, se obtienen unos datos más manejables y más significativos (tablas 24 y 25).

<i>Campos Introducidos</i>	<i>Descripción</i>
PERSONAL	Cantidad de personas (medios humanos) no especialistas destinadas al incendio
MEDIOS	Cantidad de medios mecánicos (motobombas, palas ...) destinadas al incendio

Tabla 24.

<i>Campos Eliminados</i>	<i>Descripción</i>
De PER1 a PER20	Campos destinados a los identificadores de los recursos humanos involucrados en la extinción del incendio.
De MED1 a MED20	Campos destinados a los identificadores de los recursos mecánicos involucrados en la extinción del incendio

Tabla 25.

Complejidad del algoritmo de corrección de malos diseños: $k*N*F$ donde N: número de filas de la tabla y F: número de campos que deben ser corregidos y k constante que indica el coste de acumulación de los valores de los campos a un valor único. La complejidad resulta ser lineal ya que $N \gg F \gg k$.

1.3.2. Transformación.

Se define en este caso una serie o ciclo de incendios como el conjunto de incendios que se han ido produciendo en días consecutivos, sin mediar entre ellos más de un día de diferencia.

Para su detección dentro de la base de datos de trabajo se han realizado diversas acciones que luego han tenido reflejo en el algoritmo que realiza automáticamente esta labor.

- *Ordenación de todos los incendios por su fecha de inicio.*
- *Recorrido lineal por toda la tabla* analizando la distancia en días de los inicios de los incendios que se presentan consecutivos en la tabla. Cuando la diferencia es mayor a dos días se produce un cambio de ciclo en la serie de incendios, es decir, se cierra la serie actual y se comienza una nueva.
- Mientras se va recorriendo la base de datos para determinar los ciclos se van almacenando de forma acumulativa los datos que luego se utilizarán para caracterizar los ciclos.

1.3.2.1. Análisis de los ciclos y establecimiento de una tabla donde cada registro representa a un ciclo.

Una vez limpiada toda la base de datos y eliminado el ruido se puede proceder a la obtención automática de ciclos y de sus características más relevantes.

Los datos estadísticos que se han considerado relevantes para el estudio de un ciclo y que forman parte como campos de la tabla relacional en la que se van a almacenar estos ciclos son los siguientes:

Nombre del campo	Tipo de datos	Descripción
FECHA_INI	fecha	Fecha de inicio del primer incendio del ciclo.
FECHA_FIN	fecha	Fecha de inicio del último incendio del ciclo.
NUMERO	entero	Número total de incendios (de cualquier tipo) ocurridos en el ciclo.
NUMCONATOS	entero	Número total de conatos ocurridos en el ciclo.
NUMQUEMAS	entero	Número total de quemas de matorral ocurridas en el ciclo.
NUMINCENDIOS	entero	Número total de siniestros ocurridos en el ciclo.
SUPARBTOTAL	real doble	Superficie arbolada total arrasada durante el ciclo.
SUPRASTOTAL	real doble	Superficie rasa total afectada durante el ciclo.
NUMESPEC	entero	Nº de especialistas totales que han intervenido en los incendios del ciclo (técnicos, autoridad civil y agentes forestales).
PERSONAL	entero	Nº total de medios humanos involucrados en la extinción de los incendios del ciclo.
MEDIOS	entero	Nº total de medios mecánicos involucrados en la extinción de los incendios del ciclo.

Tabla 26.

De esta forma se ha creado un algoritmo para la detección y obtención de datos acumulativos de todos los ciclos que existan en la base de datos de trabajo, para lo cual se ha utilizado como herramienta Visual C++, accediendo a la base de datos en formato MS Access '97 mediante ODBC. El algoritmo en cuestión es el siguiente:

```

// variables de acceso a las tablas de la BD
CBDIncendios m_pInc; // derivada de CRecordSet sobre la tabla INCEN9192
CBDCiclos m_pCiclos; // derivada de CRecordSet sobre la tabla Ciclos

// indicadores de las fechas de inicio y final
CTime primero;
CTime ultimo;

// indicador de número de incendios
int nNumInc;
// indicadores de los número de incendios según tipos
int nNumConatos,nNumQuemas,nNumIncendios;

// indicadores del sumatorio de superficies
double dSupArbTotal, dSupRasTotal, dSupTotal;

// indicadores de sumatorios de Personal
long dNumEspec;
int numPersonal, numMedios;

// indicador del número de ciclo
int NumCiclo = 0;

// inicialización de las Bases de Datos
m_pInc.Open();
m_pCiclos.Open();

// ordenación de los incendios mediante su fecha de inicio
m_pInc.m_strSort = "Fecha_Ini";
m_pInc.Requery();

// se deben detectar ciclos en la BD.
// variables temporales para determinar las diferencias en días.
CTime t,tant;
CTimeSpan st;

/* BUCLE DE RECORRIDO DE TODOS LOS INCENDIOS */
do {
    // inicialización de los indicadores estadísticos
        // con los valores del primer incendio del ciclo
    dSupArbTotal = m_pInc.m_SUP_ARBO;
    dSupRasTotal = m_pInc.m_SUP_RASA;
    dSupTotal      = m_pInc.m_SUP_TOTAL;
    dNumEspec      =      m_pInc.m_AGEN_FOR + m_pInc.m_AUT_CIVIL + +
    m_pInc.m_TECNICOS;
    numPersonal = m_pInc.m_Personal;
    numMedios     = m_pInc.m_Medios;

    nNumInc=1;
    // inicializacion de los datos de incendios por tipo
    nNumConatos = nNumQuemas = nNumIncendios = 0;
    if (m_pInc.m_TIPO == "C") nNumConatos++;
        else if (m_pInc.m_TIPO == "Q") nNumQuemas++;
            else if (m_pInc.m_TIPO == "I") nNumIncendios++;
    tant = m_pInc.m_FECHA_INI;
    // fecha inicial del ciclo
}

```

```

primero = tant;

/* BUCLE DE ESTUDIO DE UN CICLO */
do {

    if (m_pInc.IsEOF()) break; // si hemos llegado al final paramos
        else m_pInc.MoveNext(); // nos movemos al incendio siguiente

        // se obtiene la diferencia en días entre dos incendios consecutivos.
    t = m_pInc.m_FECHA_INI;
    st = t - tant;
    long d = st.GetDays();

        // si la diferencia es mayor que dos se ha llegado al final de un ciclo
    if (d > 2) {
        ultimo = tant;
        // se insertan los datos estadísticos del ciclo en la tabla
        Insertar(&m_pCiclos,primero,ultimo,nNumInc,NumCiclo,
                 nNumConatos, nNumQuemas, nNumIncendios,
                 dSupArbTotal,dSupRasTotal,dSupTotal,dNumEspec,
                 numPersonal,numMedios);
        break;
    }
    else {
        tant = t;
        /* actualización de datos estadísticos */
        // superficies
        dSupArbTotal += m_pInc.m_SUP_ARBO;
        dSupRasTotal += m_pInc.m_SUP_RASA;

        // personal
        dNumEspec     += m_pInc.m_AGEN_FOR + m_pInc.m_AUT_CIVIL +
        m_pInc.m_TECNICOS;
        numPersonal += m_pInc.m_Personal;
        numMedios += m_pInc.m_Medios;

        // incendios
        nNumInc++;
        // actualización de los datos de incendios según tipo
        if (m_pInc.m_TIPO == "C") nNumConatos++;
        else if (m_pInc.m_TIPO == "Q") nNumQuemas++;
            else if (m_pInc.m_TIPO == "I") nNumIncendios++;
    }

}while (1);
if (m_pInc.IsEOF())
    break; // si hemos llegado al final paramos.
else NumCiclo++; // aumentamos el número de ciclos y seguimos.

}while(1);

return;

```

Algoritmo 2: Extracción de ciclos.

Complejidad del algoritmo de detección de ciclos: $k*N$ lineal siendo N el número de incendios de la tabla y k ($k << N$) indica el coste de la adquisición de los datos característicos de los ciclos mediante acumulación.

Para llevar a cabo la inserción en una tabla de los datos obtenidos por el algoritmo anterior se utilizó la función Insertar(), que recibe como parámetros todos los datos relevantes del ciclo y los almacena en la base de datos mediante operaciones que permite el objeto CRecordSet de las MFC.

```

Void Insertar(CBDCiclos* m_pSet, CTime FechaIni, CTime FechaFin, int num, int id,
              int numC, int numQ, int numI, double dSupArb, double dSupRas, double dSupTotal,
              long numE, int numP, int numM)
{
    // operaciones iniciales de preparación de la BD
    m_pSet->AddNew();
    UpdateData(true);

    /* INSERCIÓN DE LOS DATOS EN LOS CAMPOS */
    m_pSet->m_Id = id; //ID
    m_pSet->m_Fecha_Ini = FechaIni; // FECHA_INI
    m_pSet->m_Fecha_Fin = FechaFin; // FECHA_FIN
    m_pSet->m_Numero = num; // NUMERO (total de incendios)
    m_pSet->m_NumConatos = numC; //NUMCONATOS
    m_pSet->m_NumQuemas = numQ; // NUMQUEMAS
    m_pSet->m_NumIncendios = numI; // NUMINCENDIOS
    m_pSet->m_NumEspecialistas = numE; //NUMESPEC
    m_pSet->m_SupArbTotal = dSupArb; //SUPARBTOTAL
    m_pSet->m_SupRasTotal = dSupRas; // SUPRASTOTAL
    m_pSet->m_Personal = numP; // PERSONAL
    m_pSet->m_Medios = numM; //MEDIOS

    //operaciones de manejo de la base de datos por parte de VISUAL C++
    if(m_pSet->CanUpdate()){
        m_pSet->Update();
    }

    if(!m_pSet->IsEOF()){
        m_pSet->MoveLast();
    }
    m_pSet->Requery();

    UpdateData(false);
}

```

Algoritmo 3: Función de inserción de los datos en la tabla.

Para todas las labores que siguen, se ha construido una aplicación que permite que todas las operaciones se realicen paso a paso, comprobándose cada vez los efectos que se tienen sobre la base de datos albergando todos los algoritmos y funciones auxiliares indicadas durante la exposición del preproceso y la primera transformación de los datos.

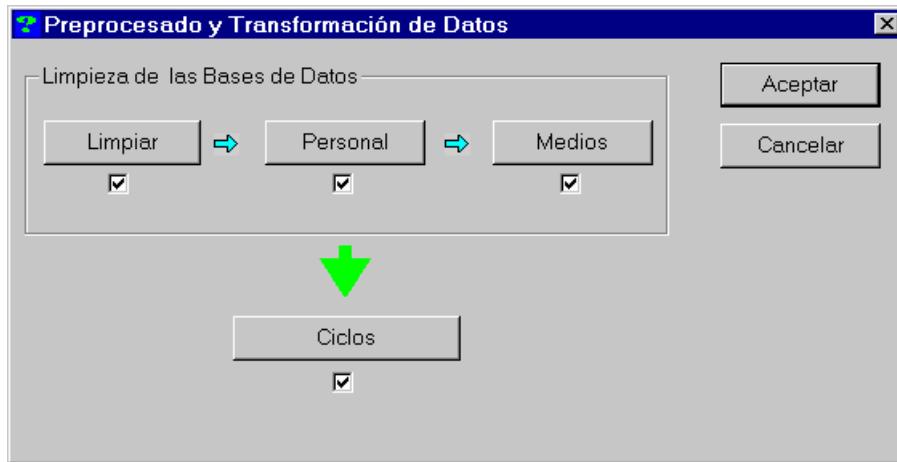


Figura 32: Interfaz de las operaciones descritas.

En este interfaz se pueden ver los diferentes botones que suponen la utilización de cada uno de los algoritmos sobre la base de datos y un control *checkbox* que les acompaña para indicar que ese proceso se ha culminado con éxito.

El resultado de este proceso han sido 37 ciclos, que quedan reflejados en la siguiente tabla:

A	B	C	D	E	F	G	H	I	J	K	L
0	2/01/91	2/01/91	1	0	0	1	0,01	0,5	3	1	1
1	15/01/91	15/01/91	1	1	0	0	0	0,001	1	1	0
2	24/01/91	30/01/91	9	2	4	3	0	0,272	10	10	2
3	23/02/91	25/02/91	12	1	8	3	0	0,395	14	13	2
4	4/03/91	4/03/91	1	0	0	1	0,008	0,03	3	1	1
5	11/03/91	11/03/91	1	0	1	0	0	0,03	1	1	0
6	20/03/91	20/03/91	3	1	0	2	0,032	0,13	2	3	1
7	29/03/91	2/04/91	45	17	19	9	0,213	1,209	31	60	45
8	9/04/91	11/04/91	20	6	10	4	0	0,406	20	23	14
9	14/04/91	24/04/91	48	24	14	10	0,392	0,824	64	71	40
10	27/04/91	27/04/91	1	1	0	0	0	0,01	0	1	1
11	8/05/91	29/05/91	113	59	18	36	0,99	3,641	162	185	126
12	1/06/91	8/06/91	7	4	2	1	0	0,076	9	7	5
13	11/06/91	19/06/91	19	15	0	4	0,169	0	24	30	30
14	22/06/91	6/07/91	57	39	10	8	0	1,841	72	102	87

15	10/07/91	10/09/91	554	375	94	85	0	14,922	744	1306	1196
16	14/09/91	25/09/91	87	58	24	5	0	1,335	94	196	198
17	2/10/91	2/10/91	1	1	0	0	0	0	0	1	0
18	6/10/91	7/10/91	8	6	2	0	0	0,109	10	20	6
19	1/11/91	1/11/91	1	0	1	0	0	0,03	1	2	0
20	24/11/91	24/11/91	1	0	1	0	0	0,02	4	2	1
21	29/11/91	30/11/91	2	1	1	0	0	0,03	2	2	0
22	5/12/91	7/12/91	16	10	5	1	0	0,396	9	25	7
23	14/12/91	16/12/91	6	1	4	1	0	0,245	7	14	3
24	21/12/91	8/01/92	213	98	78	37	63,876	434,756	147	383	136
25	16/01/92	20/01/92	9	8	1	0	1,01	2,71	7	12	4
26	23/01/92	11/02/92	424	233	112	79	206,3	1064,74	234	635	187
27	15/02/92	23/03/92	916	635	189	92	183,1	1074,52	366	1406	515
28	10/04/92	23/05/92	268	195	41	32	139,2	293,4	223	494	305
29	13/06/92	15/06/92	7	6	0	1	0,8	3,24	7	9	5
30	19/06/92	20/06/92	3	2	0	1	1,7	1,8	2	4	3
31	25/06/92	26/06/92	2	2	0	0	0	0,3	1	2	0
32	29/06/92	18/07/92	53	51	1	1	2,08	8,57	38	70	54
33	21/07/92	7/08/92	133	110	9	14	0	82,9	136	266	217
34	12/08/92	27/08/92	80	70	6	4	6,15	0	68	155	142
35	2/09/92	21/09/92	80	74	4	2	3,92	0	87	155	142
36	3/10/92	3/10/92	1	0	0	1	1	0,75	1	2	1

Tabla 26: Ciclos de incendios 1991 - 92.

Leyenda:

A: Número de Identificación.

B: Fecha de Inicio del ciclo.

C: Fecha de Finalización del ciclo.

D: Número total de sucesos.

E: Número de Conatos.

F: Número de Quemas.

G: Número de Incendios.

H: Superficie arbolada quemada.

I: Superficie rasa quemada.

J: Número de especialistas (Técnicos, Agentes Forestales ...).

K: Número de brigadas de personal.

L: Número de medios mecánicos (Terrestres, aéreos ...).

La tabla 26 muestra los 37 ciclos de incendios entre dos períodos de lluvias, que han tenido lugar en la zona estudiada durante los años 1991 y 1992. El siguiente paso será comprobar qué estructura de clases determinan los propios datos de estos ciclos (*clustering*), establecer una correlación entre los *clusters* descubiertos y los prototipos sugeridos por los expertos y, por último, encuadrar cada uno de estos ciclos en su correspondiente prototipo.

Es decir, primero se ha realizado un *clustering* sensible al contexto (detección de ciclos) y a continuación se hará uno en un espacio lingüístico inducido (se comenzará con un proceso de *clustering* sobre la tabla de ciclos y se concluirá con uno de clasificación de cada ciclo en un prototipo).

1.3.2.2. *Clustering jerárquico sobre la tabla de ciclos.*

Con el fin de detectar las relaciones entre los ciclos, para obtener aquellos de escasa, mediana y alta siniestralidad, se realiza un proceso de *clustering* jerarquizado mediante la técnica de Emparrillados (*Repertory Grids*) descrita con detalle en el Apéndice 1. El conjunto de elementos es el constituido por los 37 ciclos, y las construcciones son las 7 que se detallan a continuación:

<i>Construcción</i>	Nº DE INCENDIOS, C1		
<i>Valores</i>	Muy pocos	[0 - 20]	1
	Pocos	[21- 30]	2
	Regular	[31- 50]	3
	Nº Alto	[51 - 100]	4
	Muy alto	[101 +]	5

Tabla 27: Construcción Número total de incendios.

<i>Construcción</i>	SUPERFICIE ARB. QUEMADA, C2		
<i>Valores</i>	Muy poca	[0 - 0,2]	1
	Poca	[0,21 - 0,4]	2
	Regular	[0,41 - 1]	3
	Bastante	[1,1 - 20]	4
	Mucha	[21 +]	5

Tabla 28: Construcción Superficie arbolada quemada.

<i>Construcción</i>	SUPERFICIE RASA QUEMADA, C3
<i>Valores</i>	Muy poca [0 - 1] 1
	Poca [1,1 - 5] 2
	Regular [5,1 - 100] 3
	Bastante [101 - 1000] 4
	Mucha [1001 +] 5

Tabla 29: Construcción Superficie rasa quemada.

<i>Construcción</i>	Nº DE ESPECIALISTAS, C4
<i>Valores</i>	Muy pocos [0- 10] 1
	Pocos [11 - 20] 2
	Regular [21 - 50] 3
	Nº Alto [51 - 100] 4
	Muy alto [101 +] 5

Tabla 30: Construcción Número de especialistas.

<i>Construcción</i>	Nº DE PERSONAL, C5
<i>Valores</i>	Muy pocos [0 - 10] 1
	Pocos [11 - 50] 2
	Regular [51 - 100] 3
	Nº Alto [101 - 300] 4
	Muy alto [301 +] 5

Tabla 31: Construcción Número de brigadas de personal.

<i>Construcción</i>	Nº DE MEDIOS, C6
<i>Valores</i>	Muy pocos [0 - 10] 1
	Pocos [11 - 30] 2
	Regular [31 - 50] 3
	Nº Alto [51 - 100] 4
	Muy alto [101 +] 5

Tabla 32: Construcción Número de medios mecánicos.

<i>Construcción</i>	Nº DE DÍAS, C7
<i>Valores</i>	Muy pocos [0 - 3] 1
	Pocos [4 - 10] 2
	Regular [11 - 20] 3
	Nº Alto [21 - 30] 4
	Muy alto [31 +] 5

Tabla 33: Construcción Número de días.

Con lo que la malla de repertorio queda como se refleja en la tabla 34:

Ciclo	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Ciclo
C1	1	1	1	1	1	1	3	1	3	1	5	1	1	4	5	4	1	1	C1	
C2	1	1	1	1	1	1	2	1	2	1	3	1	1	1	1	1	1	1	C2	
C3	1	1	1	1	1	1	2	1	1	1	2	1	1	2	3	2	1	1	C3	
C4	1	1	1	2	1	1	1	3	2	4	1	5	1	3	4	5	4	1	C4	
C5	1	1	1	2	1	1	1	3	2	3	1	4	1	2	4	5	4	1	C5	
C6	1	1	1	1	1	1	3	2	3	1	5	1	2	4	5	5	1	1	C6	
C7	1	1	2	1	1	1	2	1	3	1	4	2	2	3	5	3	1	1	C7	

Ciclo	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	Ciclo
C1	1	1	1	1	1	5	1	5	5	5	1	1	1	4	5	4	4	1	C1
C2	1	1	1	1	1	5	4	5	5	5	3	4	1	4	1	4	4	3	C2
C3	1	1	1	1	1	4	2	5	5	4	2	2	1	3	3	1	1	1	C3
C4	1	1	1	1	1	5	1	5	5	5	1	1	1	3	5	4	4	1	C4
C5	1	1	1	2	2	5	2	5	5	5	1	1	1	3	4	4	4	1	C5
C6	1	1	1	1	1	5	1	5	5	5	1	1	1	4	5	5	5	1	C6
C7	1	1	1	1	1	3	2	3	5	5	1	1	1	3	3	3	3	1	C7

Tabla 34: Malla de repertorio.

Para realizar un análisis de *clusters* sobre elementos, se construye una matriz de proximidad (malla de repertorio reducida), que representa las diferentes similitudes de los elementos, una matriz de 37x37 elementos (los ciclos), que por encima de la diagonal representan las distancias entre los diferentes ciclos. Pasando estos valores a porcentajes, se crea la tabla reducida a porcentaje, con los siguientes algoritmos:

1. Para cada elemento hacer
 - 1.1. Para cada elemento siguiente a él hacer
 - 1.2. Sumar los valores absolutos de las diferencias de los valores de las distintas construcciones
 2. Estandarizar la distancia suma a conceptos de porcentaje
 3. Devolver el valor de similaridad

Algoritmo 4: Pseudocódigo del algoritmo que obtiene la tabla reducida a porcentaje.

```

int diferencia (int col1, int col2)
{
    int fila;
    int dist = 0;
    int i;
    /* Se calcula valor absoluto de la distancia entre los
    dos elementos que se introducen como parámetros */
    for (i = 0; i < FILAS; i++)
        dist += abs(MATRIZ[i][col1] - MATRIZ[i][col2]);
    /* se transforma a porcentaje
    * N = nº de construcciones (7)
    * D = diferencia máxima posible para una construcción (4)
    */
    int result = 100 - ( (100 * dist) / (D * N) );
    return result;
}

```

Algoritmo 5: Algoritmo que obtiene la tabla reducida a porcentaje.

1. Tomar como parámetros los elementos más similares
2. Recalcular distancias
 - 2.1 Ignorar las filas y columnas que correspondan a los más similares
 - 2.2 Si uno de los elementos a comparar ha sido elegido como más similar
 - 2.2.1 Se toma la similaridad con el grupo recién obtenido, o sea, el mínimo de similaridad anterior de los elementos del grupo con el otro elemento
 - 2.3 Si no
 - 2.3.1 Se mantiene la similaridad anteriormente calculada
3. Se recomponen los descriptores de los elementos agrupando los que son más similares

Algoritmo 6: Pseudocódigo del algoritmo de readaptación de la matriz.

```

CMatrix *CMatrix::Eliminar(int fila1, int fila2)
{
    CMatrix *T = new CMatrix(m_nFilas-1,m_nColumnas-1,szDesc);
    int i,j;
    // recalcular distancias
    for (i = 0; i < m_nFilas; i++) {
        for (j = i+1; j < m_nColumnas; j++) {
            // eliminar y columnas de los elementos más similares
            if (j == fila2 || i == fila2) continue;
            if (i == fila1) {
                /* tomar como similitud al grupo de similares el mínimo
                 * de los dos elementos con el elemento externo
                */
                int dx = (j > fila2)?j-1:j;
                int aux = (j < fila2) ? m_Matriz[j][fila2]:m_Matriz[fila2][j];
                int valor = (m_Matriz[i][j] <= aux)? m_Matriz[i][j] : aux;
                T->PonerValor(i,dx,valor);
            }
            else {
                /* mantener los valores entre los elementos no indicados
                 * como
                 * similares
                */
                int dxCol = (j > fila2)?j-1:j;
                int dxFil = (i > fila2)?i-1:i;
                T->PonerValor(dxFil,dxCol,m_Matriz[i][j]);
            }
        }
    }

    // recomponer descriptores
    CambiarDescriptor(fila1,fila2);

    return T;
}

```

Algoritmo 7: Algoritmo de readaptación de la matriz.

1. Se construye con la función de similitud una matriz de proximidad que es cuadrada y donde cada elemento de ella mide la similitud entre el elemento fila y columna que presenta.
2. Repetir mientras se encuentren similares
 - 2.1 Se toman los elementos que tienen más similaridad
 - 2.2 Se eliminan la fila y la columna
 - 2.3 Se agrupan los elementos más similares en la fila de uno de ellos
 - 2.4 Se recalculan todas las distancias
 - 2.4.1 La similitud entre elementos que no han sido elegidos como más similares.
 - 2.4.2 La similitud de los elementos con el nuevo grupo de similares se calcula con el mínimo de la similitud de ese elemento con los demás elementos del grupo.
3. Devolver Resultados

Algoritmo 8: Pseudocódigo del algoritmo de cálculo del dendograma.

```

/* FASE 1 */
// se construye una matriz de proximidad que es un emparrillado reducido
Cmatriz *M = new Cmatriz(COLUMNAS,COLUMNAS);

for (j = 0; j < NumColumnas; j++) {
    int h;
    for ( h = j+1; h < NumColumnas; h++) {
        // se coloca el valor de la similaridad porcentual entre elementos
        int dif = diferencia(j,h);
        M->PonerValor(j,h,dif);
    }
}

/* FASE 2 */
while (!M->EsFinal()) {
    int max1,max2;
    // 2.1 calcular los elementos de similaridad máxima
    // y regresar u grado de similaridad
    int similaridad = M->Maximo(&max1,&max2);
    Cmatriz *Aux = M;
    /* 2.2 eliminar la fila y columna y
     * calcular la nueva matriz de similaridad
     */
    M = M->Eliminar(max1,max2);
    delete Aux;
}

```

Algoritmo 9: Algoritmo de cálculo del dendograma.

Como resultado se obtienen los siguiente porcentajes de similaridad entre elementos y sus agrupaciones:

(0,1) -> 100%
((0,1),4) -> 100%
(((0,1),4),5) -> 100%
((((0,1),4),5),6) -> 100%
((((((0,1),4),5),6),10) -> 100%
((((((0,1),4),5),6),10),17) -> 100%
((((((0,1),4),5),6),10),17),19) -> 100%
((((((0,1),4),5),6),10),17),19),20) -> 100%
((((((0,1),4),5),6),10),17),19),20),21) -> 100%
((((((0,1),4),5),6),10),17),19),20),21),31) -> 100% (**A**)

(2,12) -> 100% (**B**)

(18,22) -> 100%

((18,22),23) -> 100% (**C**)

(34,35) -> 100% (**D**)

((((((((0,1),4),5),6),10),17),19),20),21),31),(2,12)) -> 97%

(3,8) -> 97%

(14,16) -> 97%

(24,26) -> 97%

(27,28) -> 97%

(29,30) -> 97%

((((((((0,1),4),5),6),10),17),19),20),21),31),(2,12)),((18,22),23)) -> 93%

((29,30),36) -> 93%

((((((((0,1),4),5),6),10),17),19),20),21),31),(2,12)),((18,22),23)),(3,8)) -> 90%

(7,9) -> 90%

(15,33) -> 90%

((24,26),(27,28)) -> 90%

(25,((29,30),36)) -> 90%

((((((((0,1),4),5),6),10),17),19),20),21),31),(2,12)),((18,22),23)),(3,8)),13) -> 83%

(11,(15,33)) -> 83%

((14,16),(34,35)) -> 83%

((7,9),((14,16),(34,35))) -> 79%

((((((((0,1),4),5),6),10),17),19),20),21),31),(2,12)),((18,22),23)),(3,8)),13),(25,((29,30),36)))
-> 75%

((7,9),((14,16),(34,35))),32) -> 75%

((11,(15,33)),((24,26),(27,28))) -> 75%

((((7,9),((14,16),(34,35))),32),((11,(15,33)),((24,26),(27,28)))) -> 61%

((((((((0,1),4),5),6),10),17),19),20),21),31),(2,12)),((18,22),23)),(3,8)),13),(25,((29,30),36)))
(((7,9),((14,16),(34,35))),32),((11,(15,33)),((24,26),(27,28)))) -> 40%

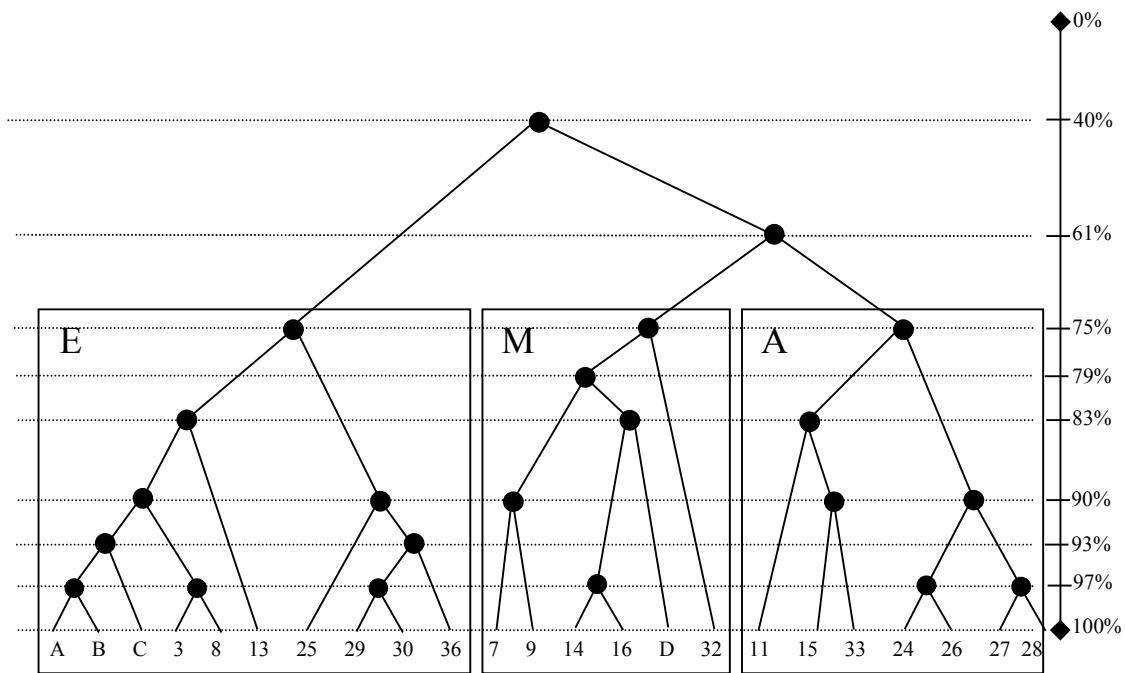


Figura 33: Dendrograma final del proceso de *clustering* jerárquico.

E: Siniestralidad escasamente progresiva.

M: Siniestralidad medianamente progresiva.

A: Siniestralidad altamente progresiva.

1.3.2.3. *Ordenación de los ciclos según la evolución de su siniestralidad.*

Una vez realizado este proceso de *clustering* sobre los ciclos, se está en condiciones de ordenarlos en función de su siniestralidad. Para ello, realizados test sobre varios expertos, se ha llegado a la conclusión de que la evolución de la siniestralidad está básicamente vinculada al numero de siniestros ocurridos (sin distinguir entre incendios, superficie arbolada mayor a media hectárea, conatos, superficie total menor a media hectárea y quemas, superficie rasa mayor que media hectárea), porque cualquiera de los siniestros, por pequeño que haya sido, podría haber tenido consecuencias peores de no haber actuado a tiempo. El otro factor influyente en la evolución de la siniestralidad es la superficie que se ha quemado, dándole un mayor peso a la superficie arbolada.

Teniendo en cuenta estos factores, se está en condiciones de definir una medida Heurística de siniestralidad del ciclo, expresada del siguiente modo:

Número total de incendios/100

+

$$[(1 * \text{Superficie arbolada total}) + (0,5 * \text{Superficie rasa total})]/100$$

Presentados casos de ejemplo, los expertos concuerdan en que un ciclo sería de siniestralidad escasamente progresiva cuando esta medida estuviera por debajo de 0,4, entre este valor y 1 sería un ciclo de siniestralidad medianamente progresiva y a partir de 1 la siniestralidad sería altamente progresiva.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	P
17	2/10/91	2/10/91	1	1	0	0	0	0	0	1	0	1	0,01	E
1	15/01/91	15/01/91	1	1	0	0	0	0,001	1	1	0	1	0,0105	E
10	27/04/91	27/04/91	1	1	0	0	0	0,01	0	1	1	1	0,0105	E
20	24/11/91	24/11/91	1	0	1	0	0	0,02	4	2	1	1	0,0101	E
5	11/03/91	11/03/91	1	0	1	0	0	0,03	1	1	0	1	0,0105	E
19	1/11/91	1/11/91	1	0	1	0	0	0,03	1	2	0	1	0,0105	E
4	4/03/91	4/03/91	1	0	0	1	0,008	0,03	3	1	1	1	0,0103	E
0	2/01/91	2/01/91	1	0	0	1	0,01	0,5	3	1	1	1	0,0126	E
21	29/11/91	30/11/91	2	1	1	0	0	0,03	2	2	0	2	0,0205	E
31	25/06/92	26/06/92	2	2	0	0	0	0,3	1	2	0	2	0,0215	E
36	3/10/92	3/10/92	1	0	0	1	1	0,75	1	2	1	1	0,0235	E
6	20/03/91	20/03/91	3	1	0	2	0,032	0,13	2	3	1	1	0,0307	E
30	19/06/92	20/06/92	3	2	0	1	1,7	1,8	2	4	3	2	0,056	E
23	14/12/91	16/12/91	6	1	4	1	0	0,245	7	14	3	3	0,0625	E
12	1/06/91	8/06/91	7	4	2	1	0	0,076	9	7	5	8	0,0708	E
18	6/10/91	7/10/91	8	6	2	0	0	0,109	10	20	6	2	0,0805	E
2	24/01/91	30/01/91	9	2	4	3	0	0,272	10	10	2	7	0,0913	E
29	13/06/92	15/06/92	7	6	0	1	0,8	3,24	7	9	5	3	0,0942	E
25	16/01/92	20/01/92	9	8	1	0	1,01	2,71	7	12	4	5	0,1136	E
3	23/02/91	25/02/91	12	1	8	3	0	0,395	14	13	2	3	0,1219	E
22	5/12/91	7/12/91	16	10	5	1	0	0,396	9	25	7	3	0,1619	E
13	11/06/91	19/06/91	19	15	0	4	0,169	0	24	30	30	9	0,1916	E
8	9/04/91	11/04/91	20	6	10	4	0	0,406	20	23	14	3	0,2020	E
7	29/03/91	2/04/91	45	17	19	9	0,213	1,209	31	60	45	4	0,4581	M
9	14/04/91	24/04/91	48	24	14	10	0,392	0,824	64	71	40	11	0,4880	M
14	22/06/91	6/07/91	57	39	10	8	0	1,841	72	102	87	15	0,5792	M
32	29/06/92	18/07/92	53	51	1	1	2,08	8,57	38	70	54	20	0,5936	M
35	2/09/92	21/09/92	80	74	4	2	3,92	0	87	155	142	20	0,8392	M
34	12/08/92	27/08/92	80	70	6	4	6,15	0	68	155	142	16	0,8615	M
16	14/09/91	25/09/91	87	58	24	5	0	1,335	94	196	198	12	0,8766	M
11	8/05/91	29/05/91	113	59	18	36	0,99	3,641	162	185	126	22	1,1581	A
33	21/07/92	7/08/92	133	110	9	14	0	82,9	136	266	217	17	1,7445	A
24	21/12/91	8/01/92	213	98	78	37	63,876	434,756	147	383	136	18	4,9425	A
28	10/04/92	23/05/92	268	195	41	32	139,2	293,4	223	494	305	44	5,539	A
15	10/07/91	10/09/91	554	375	94	85	0	14,922	744	1306	1196	61	5,6146	A

26	23/01/92	11/02/92	424	233	112	79	206,3	1064,74	234	635	187	19	11,626	A
27	15/02/92	23/03/92	916	635	189	92	183,1	1074,52	366	1406	515	39	16,363	A

Tabla 35: Ordenación de los ciclos según su peligrosidad.

Leyenda:

A: Número de Identificación.

B: Fecha de Inicio del ciclo.

C: Fecha de Finalización del ciclo.

D: Número total de sucesos.

E: Número de Conatos.

F: Número de Quemas.

G: Número de Incendios.

H: Superficie arbolada quemada.

I: Superficie rasa quemada.

J: Número de especialistas (Técnicos, Agentes Forestales ...).

K: Número de brigadas de personal.

L: Número de medios (Terrestres, aéreos ...).

M: Número de días de duración del ciclo.

N: Valor de la medida Heurística de siniestralidad del ciclo.

P: Prototipo de evolución de la siniestralidad: E (Escasamente progresiva), M (Medianamente progresiva), A (Altamente progresiva).

Complejidad del algoritmo de clasificación y medición de ciclos: $k*N$ siendo N el número de ciclos detectados y k el coste de la métrica calculada y la asignación del valor.

1.3.3. Data Mining.

En esta parte final del proceso de Adquisición de Conocimiento, lo que se pretende es aplicar funciones de resumen sobre los ciclos de cada tipo, con el fin

de obtener los tres prototipos de evolución de la siniestralidad, para poder evaluar casos reales y llegar a predecir el comportamiento y las necesidades en los siniestros de días sucesivos.

Se presenta cada ciclo con los valores relevantes de cada uno de los días y con un gráfico que representa la evolución de la siniestralidad en el tiempo (en días), tomando como dato la media de todos los valores de cada día (ocurrencia diaria), normalizados en el intervalo [0, 10]. El patrón de este gráfico sería el presentado al principio de este proceso (fig. 34):

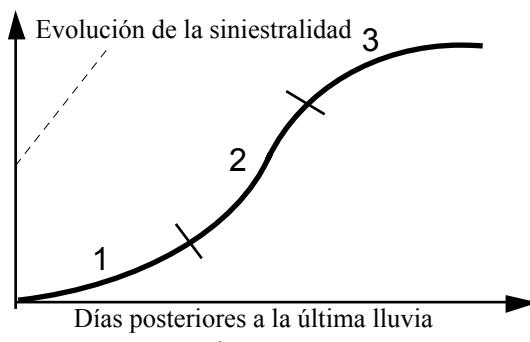


Figura 34.

- El *primer sector* representa un lento crecimiento de la siniestralidad en los días inmediatamente posteriores al periodo de lluvias. Durará desde el inicio del ciclo hasta que la ocurrencia diaria alcance el valor 4.
- El *segundo sector* expresa un alto crecimiento de la siniestralidad, especialmente en cuanto al número de incendios. Será desde que la ocurrencia diaria haya alcanzado el valor 4 por primera vez, hasta que alcance el valor 7.
- En el *tercero* se quiere representar una estabilización en cuanto al número, pero un progresivo aumento en la peligrosidad de los fuegos. Los valores irán desde que la ocurrencia diaria haya alcanzado el valor 7 por primera vez, hasta el final.

1.3.3.1. Siniestralidad escasamente progresiva.

Los ciclos 0, 1, 4, 5, 6, 10, 17, 18, 19, 20, 21, 23, 29, 30, 31 y 36 no se tienen en cuenta por ser de uno o dos días con incendios y por lo tanto, ser nula su representatividad. Teniendo esto en cuenta, los ciclos que se analizan son los

siguientes (en el eje X se representan los días del ciclo y en Y el valor de la Ocurrencia Diaria):

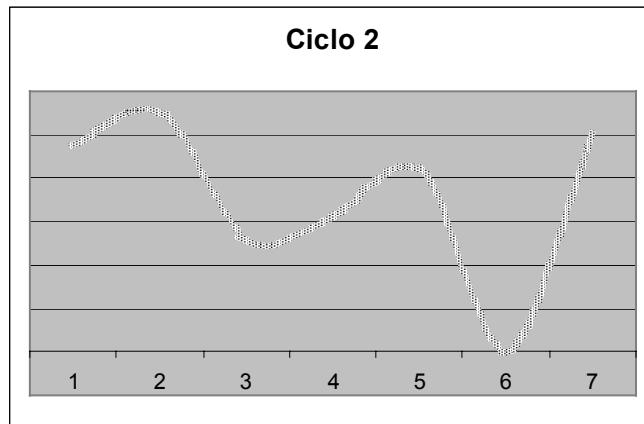


Tabla 36: Ciclo 2.

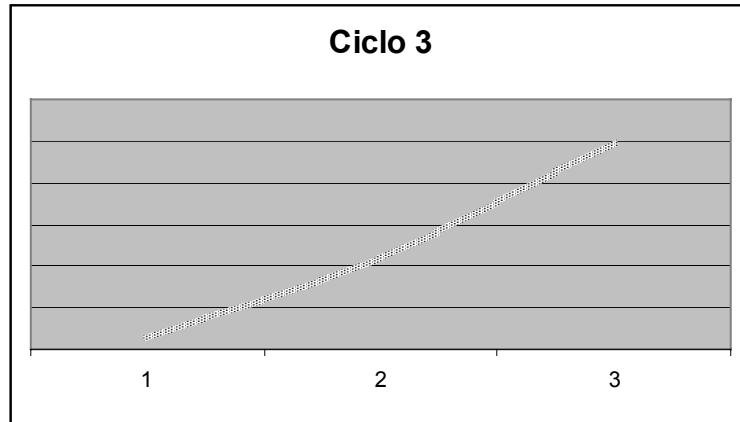
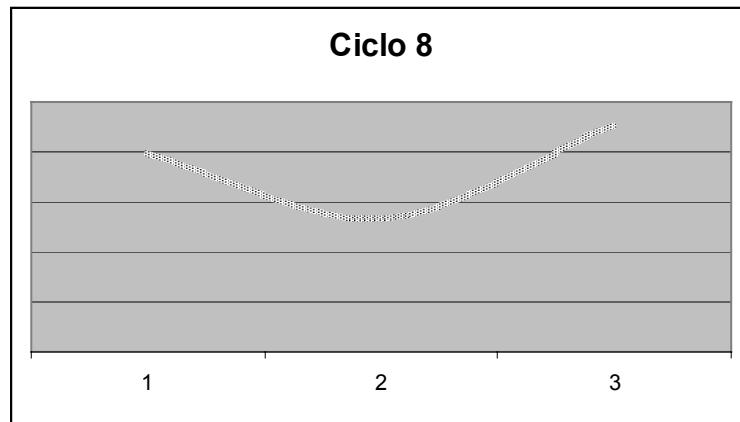
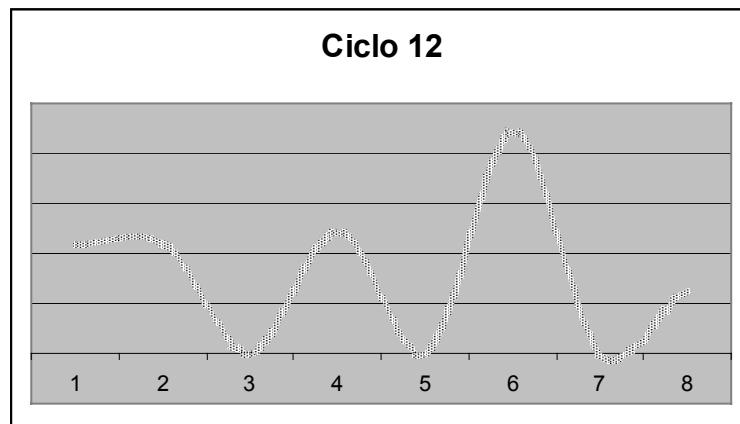


Tabla 37: Ciclo 3.



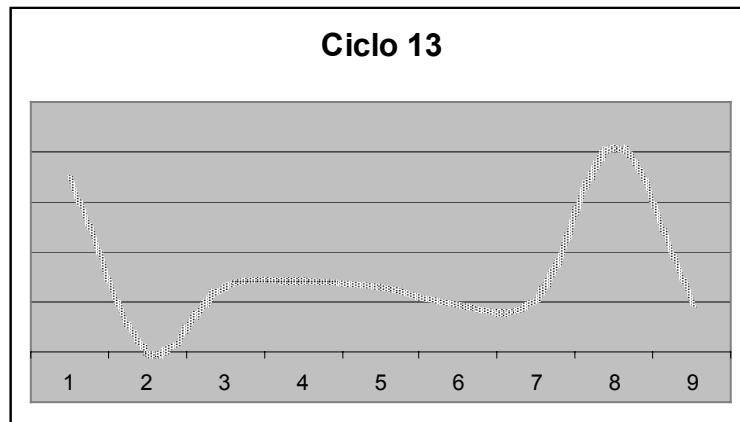
Día	NºInc	SupArb	SupRasa	Esp	Brig	MedMec	Oc.D.	Sector
09-abr-91	9	0,03	0,151	10	10	4	5,5	2
10-abr-91	5	0,01	0,095	6	5	5	3,5	2
11-abr-91	6	0,128	0,16	12	8	5	5,2	2

Tabla 38: Ciclo 8.



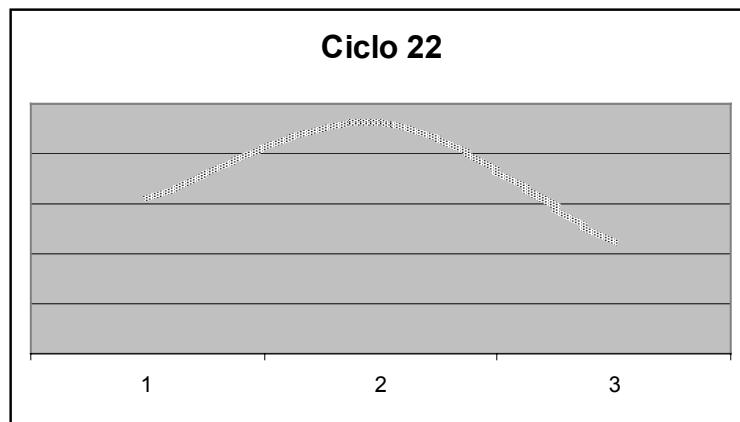
Día	NºInc	SupArb	SupRasa	Esp	Brig	MedMec	Oc.D.	Sector
01-jun-91	1	0	0,03	1	1	1	4,4	2
02-jun-91	1	0	0,03	1	1	1	4,4	2
03-jun-91	0	0	0	0	0	0	0,0	2
04-jun-91	2	0,001	0,006	1	2	1	4,9	2
05-jun-91	0	0	0	0	0	0	0,0	2
06-jun-91	2	0,01	0,01	7	2	2	8,9	3
07-jun-91	0	0	0	0	0	0	0,0	3
08-jun-91	1	0,001	0	3	1	0	2,5	3

Tabla 39: Ciclo 12.



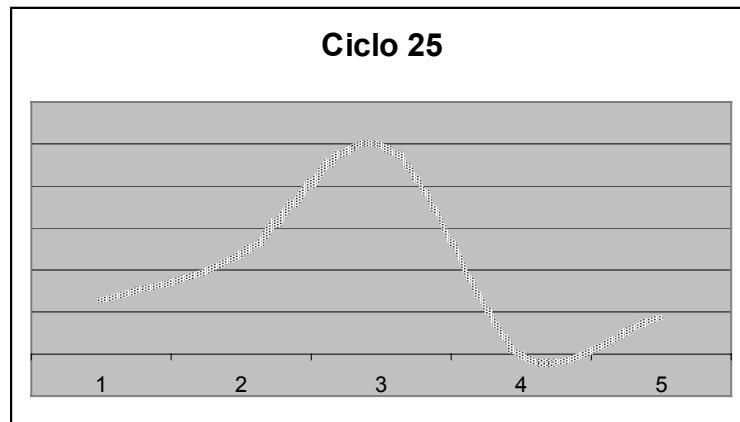
Día	NºInc	SupArb	SupRasa	Esp	Brig	MedMec	Oc.D.	Sector
11-jun-91	3	0,083	0,08	6	4	5	6,9	2
12-jun-91	0	0	0	0	0	0	0,0	2
13-jun-91	1	0,02	0,03	4	2	2	2,6	2
14-jun-91	2	0,01	0,001	4	5	2	2,9	2
15-jun-91	3	0	0,009	1	3	3	2,6	2
16-jun-91	2	0,001	0,002	1	3	2	1,9	2
17-jun-91	2	0,005	0	1	3	3	2,1	2
18-jun-91	4	0,05	0,028	14	8	11	8,3	3
19-jun-91	2	0	0,006	2	2	2	1,9	3

Tabla 40: Ciclo 13.



Día	NºInc	SupArb	SupRasa	Esp	Brig	MedMec	Oc.D.	Sector
05-dic-91	4	0	0,21	5	9	2	6,2	2
06-dic-91	6	0,01	0,116	9	9	4	9,3	3
07-dic-91	6	0	0,07	3	7	1	4,5	3

Tabla 41: Ciclo 22.



Día	NºInc	SupArb	SupRasa	Esp	Brig	MedMec	Oc.D.	Sector
16-ene-92	1	0	0,5	1	2	1	2,6	1
17-ene-92	2	0,51	0,01	2	2	1	4,8	2
18-ene-92	5	0,5	2	3	7	2	10,0	3
19-ene-92	0	0	0	0	0	0	0,0	3
20-ene-92	1	0	0,2	2	1	0	1,8	3

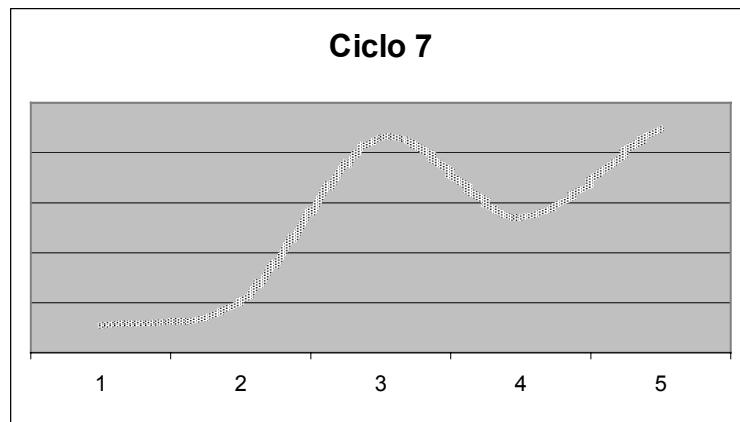
Tabla 42: Ciclo 25.

Analizando estos ciclos y aplicando funciones que los resuman, se llega a la definición del prototipo “**Siniestralidad Escasamente Progresiva**”, que es el que refleja la tabla 43:

	Sector 1	Sector 2	Sector 3
Media de días:	1	4	2
Media de incendios/día:	1	2	4
Mínimo de inc/día:	1	1	1
Máximo de inc/día:	1	4	9
Nº especialistas/día:	1	2	6
Nº de brigadas/día:	2	2	5
Nº de medios/día:	1	1	3

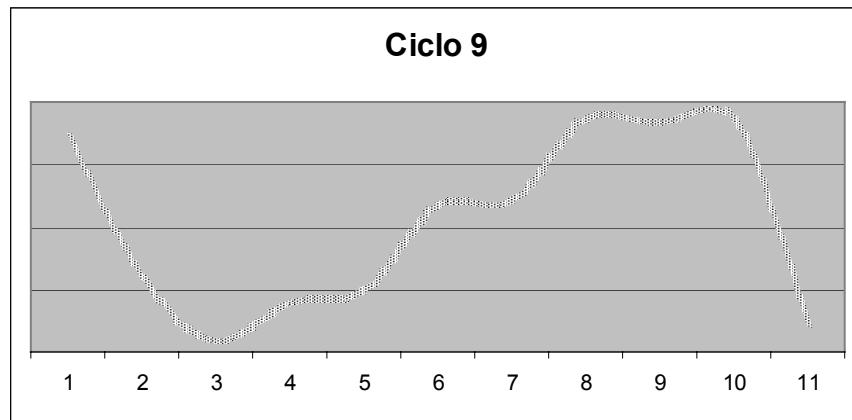
Tabla 43: Prototipo “Siniestralidad Escasamente Progresiva”.

1.3.3.2. *Siniestralidad medianamente progresiva.*



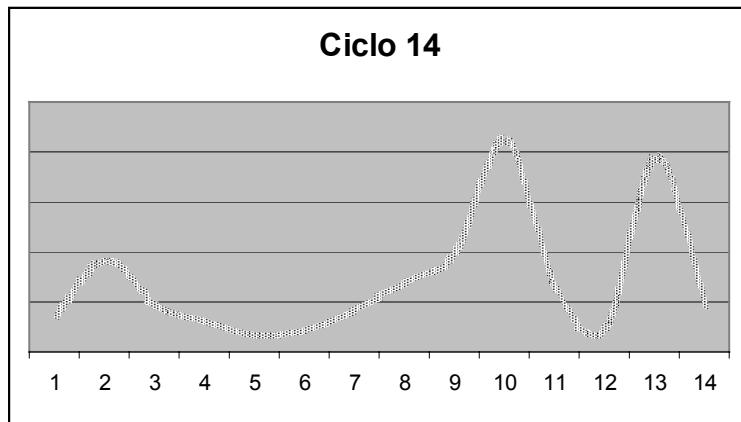
Día	NºInc	SupArb	SupRasa	Esp	Brig	MedMec	Oc.D.	Sector
29-mar-91	3	0	0,1	1	3	1	1,1	1
30-mar-91	4	0,035	0,04	2	4	3	2,1	1
31-mar-91	14	0,079	0,569	11	19	12	8,6	3
01-abr-91	11	0,011	0,23	12	12	11	5,5	3
02-abr-91	13	0,088	0,27	15	22	18	9,0	3

Tabla 44: Ciclo 7.



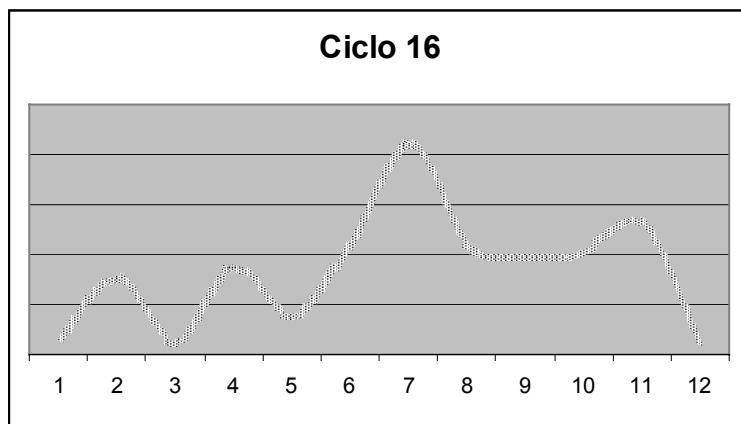
Día	NºInc	SupArb	SupRasa	Esp	Brig	MedMec	Oc.D.	Sector
14-abr-91	4	0,005	0,235	18	8	8	6,9	2
15-abr-91	4	0,005	0,013	0	5	3	2,4	2
16-abr-91	1	0	0,005	0	1	0	0,4	2
17-abr-91	2	0,01	0,005	3	4	1	1,6	2
18-abr-91	2	0,012	0,013	4	4	2	2,0	2
19-abr-91	6	0,04	0,117	6	9	2	4,8	2
20-abr-91	6	0,013	0,07	10	9	4	4,9	2
21-abr-91	7	0,158	0,16	13	9	4	7,5	3
22-abr-91	7	0,074	0,036	21	10	8	7,3	3
23-abr-91	8	0,075	0,11	11	11	8	7,4	3
24-abr-91	1	0	0,06	1	1	0	0,9	3

Tabla 45: Ciclo 9.



Día	NºInc	SupArb	SupRasa	Esp	Brig	MedMec	Oc.D.	Sector
22-jun-91	2	0,015	0,01	4	5	2	1,5	1
23-jun-91	6	0,01	0,053	8	12	9	3,7	1
24-jun-91	4	0	0,016	4	5	5	1,9	1
25-jun-91	2	0	0,035	6	2	3	1,3	1
26-jun-91	1	0	0,01	1	3	2	0,7	1
27-jun-91	3	0,003	0,003	0	3	0	0,9	1
28-jun-91	3	0	0,124	3	5	3	1,7	1
29-jun-91	6	0,009	0,024	7	7	5	2,8	1
30-jun-91	5	0	0,154	14	11	10	4,0	2
01-jul-91	6	0,124	0,505	26	19	15	8,5	3
02-jul-91	5	0	0,051	5	5	8	2,5	3
03-jul-91	2	0	0,006	2	3	2	0,9	3
04-jul-91	9	0,017	0,813	16	17	19	7,7	3
06-jul-91	3	0	0,037	7	5	4	1,9	3

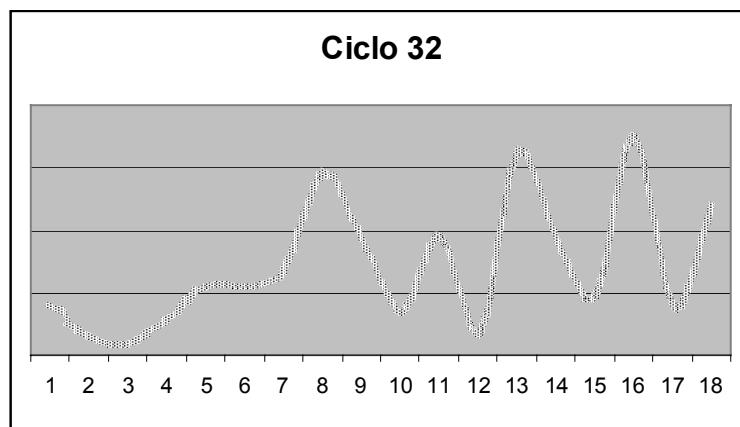
Tabla 46: Ciclo 14.



Día	NºInc	SupArb	SupRasa	Esp	Brig	MedMec	Oc.D.	Sector
14-sep-91	2	0	0,006	3	3	4	0,6	1
15-sep-91	5	0,003	0,093	13	15	20	3,1	1
16-sep-91	2	0	0,01	1	3	3	0,5	1
17-sep-91	8	0,002	0,123	8	20	23	3,5	1
18-sep-91	4	0,008	0,035	6	7	5	1,5	1
19-sep-91	11	0,012	0,068	18	22	22	4,3	2

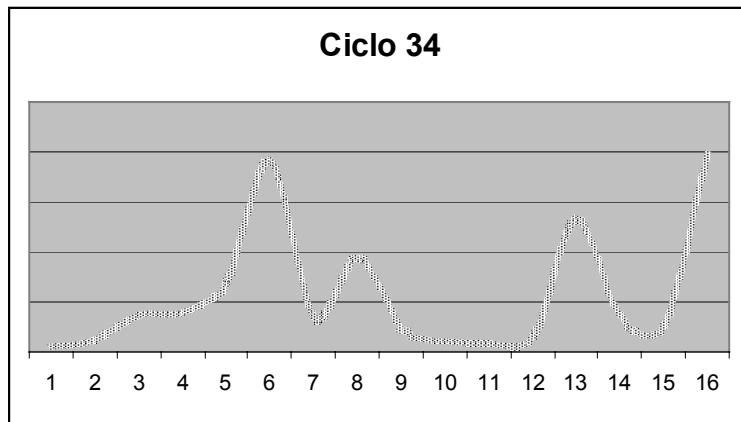
20-sep-91	19	0,01	0,286	31	43	42	8,4	3
21-sep-91	6	0,06	0,12	12	14	10	4,4	3
22-sep-91	10	0,001	0,155	13	20	18	3,9	3
23-sep-91	11	0	0,102	13	25	23	4,1	3
24-sep-91	8	0,015	0,322	12	22	25	5,3	3
25-sep-91	1	0	0,015	2	2	3	0,5	3

Tabla 47: Ciclo 16.



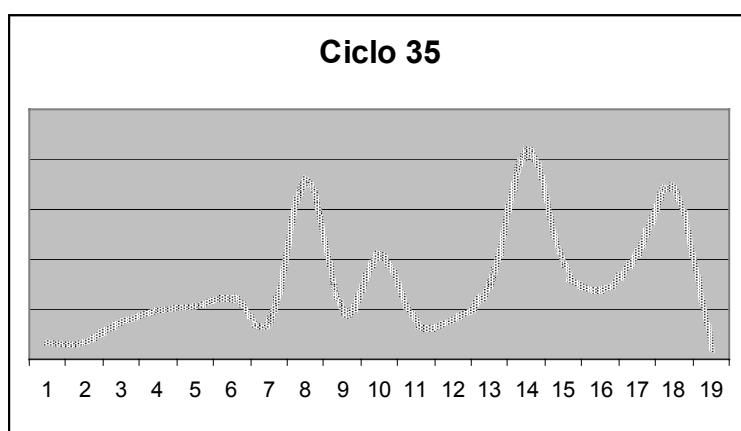
Día	NºInc	SupArb	SupRasa	Esp	Brig	MedMec	Oc.D.	Sector
29-jun-92	1	0	1	1	1	1	1,7	1
01-jul-92	1	0	0,1	0	1	1	0,6	1
02-jul-92	1	0	0	0	1	0	0,4	1
03-jul-92	2	0	0,11	1	2	1	1,1	1
04-jul-92	2	0	0,51	5	2	2	2,2	1
05-jul-92	4	0	0,4	0	4	2	2,2	1
06-jul-92	3	0	0,14	7	3	3	2,7	1
07-jul-92	4	0,4	1,3	7	6	7	5,8	2
09-jul-92	1	1,2	0	4	5	3	3,7	2
10-jul-92	3	0	0,13	0	3	1	1,4	2
11-jul-92	4	0,25	0,19	6	6	4	3,8	2
12-jul-92	1	0	0,04	1	1	1	0,7	2
13-jul-92	4	0	1,22	13	7	9	6,5	2
14-jul-92	5	0,13	0,18	4	6	4	3,6	2
15-jul-92	3	0	0,34	1	3	2	1,9	2
16-jul-92	8	0,1	1,14	8	11	7	7,0	3
17-jul-92	3	0	0,13	1	3	1	1,5	3
18-jul-92	3	0	1,64	6	5	5	4,7	3

Tabla 48: Ciclo 32.



Día	NºInc	SupArb	SupRasa	Esp	Brig	MedMec	Oc.D.	Sector
12-agosto-92	1	0,01	0	1	1	0	0,2	1
13-agosto-92	2	0	0,16	1	2	2	0,5	1
14-agosto-92	6	0,01	0,64	3	8	6	1,6	1
15-agosto-92	4	0	1,57	6	8	4	1,6	1
16-agosto-92	8	0	1,17	9	14	11	2,7	1
17-agosto-92	17	0,25	9,17	27	25	39	7,6	3
18-agosto-92	4	0	0,73	4	6	7	1,4	3
19-agosto-92	7	0,85	1,75	16	16	16	3,7	3
20-agosto-92	2	0,03	0,65	5	3	4	0,9	3
21-agosto-92	2	0	0,01	2	2	1	0,5	3
22-agosto-92	2	0	0,24	0	2	1	0,4	3
23-agosto-92	3	0	0,16	0	3	3	0,6	3
24-agosto-92	9	0	6,91	18	26	23	5,3	3
25-agosto-92	3	0	0,39	8	6	6	1,4	3
26-agosto-92	3	0	0,26	7	3	3	1,1	3
27-agosto-92	7	5	13,65	24	30	16	7,9	3

Tabla 49: Ciclo 34.



Día	NºInc	SupArb	SupRasa	Esp	Brig	MedMec	Oc.D.	Sector
02-sept-92	1	0,1	0	3	2	1	0,7	1
04-sept-92	1	0	0,2	4	1	2	0,8	1
05-sept-92	3	0,04	0,54	1	6	4	1,5	1
06-sept-92	5	0	0,1	2	8	5	2,0	1

07-sep-92	3	0,28	0,12	6	9	4	2,2	1
08-sep-92	3	0,13	0,06	15	6	4	2,5	1
09-sep-92	3	0,15	0,05	3	5	4	1,5	1
10-sep-92	9	0	5,41	14	20	21	7,1	3
11-sep-92	2	0,5	1	4	4	3	1,9	3
12-sep-92	6	0	0,86	20	10	9	4,2	3
13-sep-92	3	0,02	0,15	3	4	6	1,5	3
14-sep-92	4	0	0,25	2	5	5	1,6	3
15-sep-92	4	0	2,99	7	7	6	3,0	3
16-sep-92	10	1,87	1,34	22	24	16	8,3	3
17-sep-92	4	0,7	1,25	5	10	13	3,7	3
18-sep-92	4	0,02	1,02	10	8	7	2,8	3
19-sep-92	6	0	1,11	14	12	13	4,2	3
20-sep-92	8	0,11	6,14	18	13	18	6,8	3
21-sep-92	1	0	0,15	1	1	1	0,4	3

Tabla 50: Ciclo 35.

Analizando estos ciclos y aplicando funciones que los resuman, se llega a la definición del prototipo “**Siniestralidad Medianamente Progresiva**”, que es el que refleja la tabla 51:

	Sector 1	Sector 2	Sector 3
Media de días:	6	2	6
Media de incendios/día:	3	4	6
Mínimo de inc/día:	1	1	1
Máximo de inc/día:	8	11	19
Nº especialistas/día:	4	6	11
Nº de brigadas/día:	5	6	12
Nº de medios/día:	4	5	11

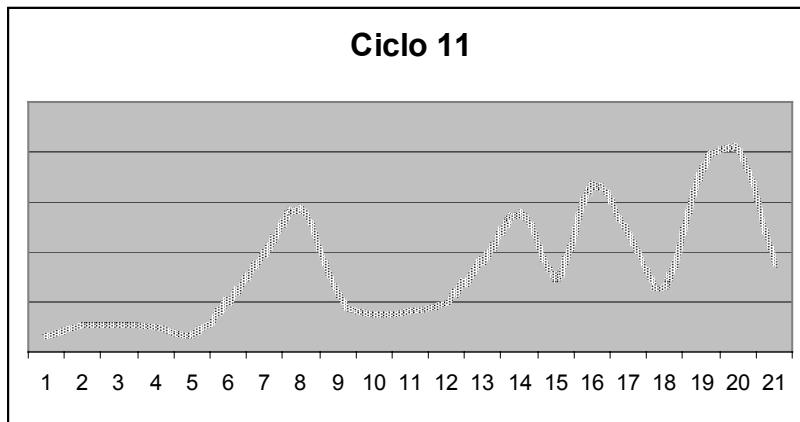
Tabla 51: Prototipo “Siniestralidad Medianamente Progresiva”.

1.3.3.3. Siniestralidad altamente progresiva.

El *primer sector* durará desde el inicio del ciclo hasta que la ocurrencia diaria alcance el valor 4.

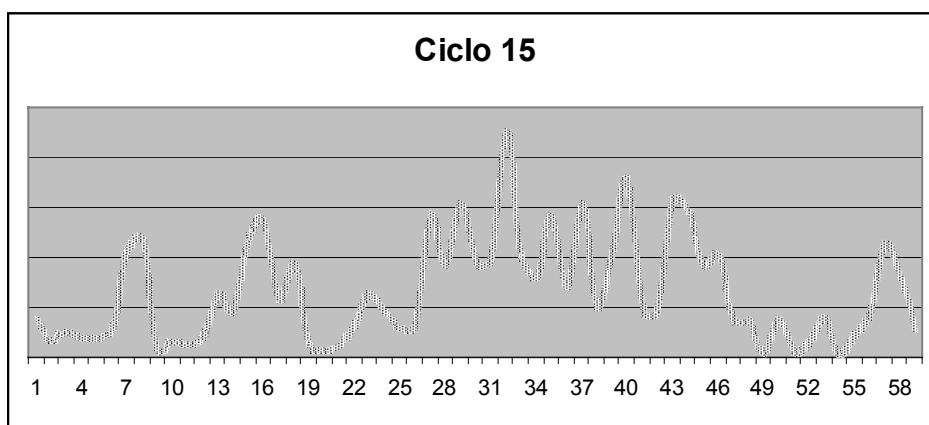
El *segundo sector* será desde que la ocurrencia diaria haya alcanzado el valor 4 por primera vez, hasta que alcance el valor 7.

En el *tercero* los valores irán desde que la ocurrencia diaria haya alcanzado el valor 7 por primera vez, hasta el final.



Día	NºInc	SupArb	SupRasa	Esp	Brig	MedMec	Oc.D.	Sector
08-may-91	1	0	0,001	5	2		1	0,6 1
09-may-91	1	0,01	0,09	6	3		2	1,1 1
10-may-91	1	0,05	0,01	5	2		1	1,1 1
12-may-91	3	0,007	0,01	1	3		3	1,0 1
13-may-91	2	0,002	0,007	3	2		1	0,7 1
14-may-91	5	0,002	0,032	6	7		5	2,0 1
15-may-91	6	0,03	0,33	13	10		10	3,9 1
16-may-91	7	0,176	0,097	16	15		10	5,6 2
17-may-91	4	0,002	0,265	7	6		5	2,2 2
18-may-91	4	0,004	0,011	5	6		2	1,5 2
19-may-91	3	0,03	0,212	4	5		1	1,7 2
20-may-91	5	0,01	0,025	7	6		4	2,0 2
21-may-91	8	0,03	0,28	12	11		4	3,6 2
22-may-91	8	0,052	0,506	18	17		10	5,5 2
23-may-91	5	0,044	0,025	19	7		4	3,0 2
24-may-91	7	0,152	0,275	33	16		10	6,6 2
25-may-91	8	0,106	0,078	17	10		9	4,6 2
26-may-91	7	0,013	0,039	7	8		6	2,6 2
27-may-91	9	0,096	1,063	23	14		15	7,4 3
28-may-91	13	0,16	0,155	25	21		18	8,0 3
29-may-91	6	0,014	0,13	16	14		5	3,5 3

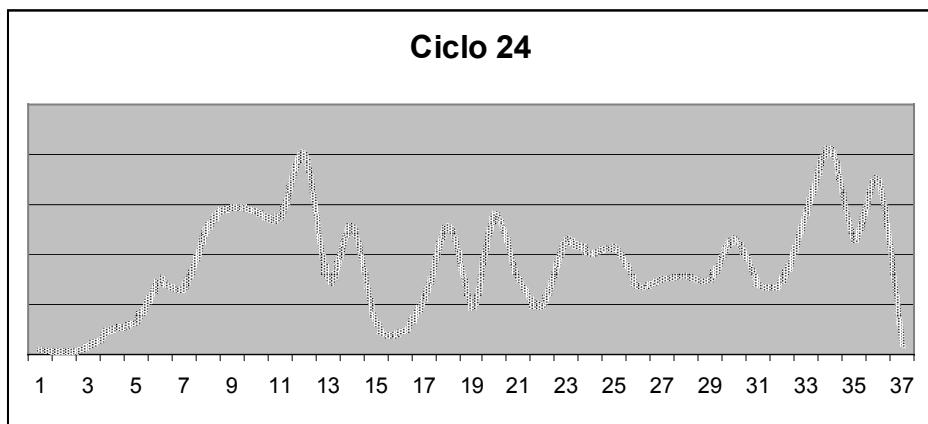
Tabla 52: Ciclo 11.



Día	NºInc	SupArb	SupRasa	Esp	Brig	MedMec	Oc.D.	Sector
10-jul-91	8	0,001	0,039	14	8	7	1,4	1
12-jul-91	3	0,021	0,081	7	4	5	0,7	1
13-jul-91	4	0,022	0,04	9	6	9	1,0	1
14-jul-91	2	0,05	0,005	6	7	7	0,8	1
15-jul-91	3	0	0,03	5	6	9	0,8	1
16-jul-91	4	0,004	0,04	10	11	9	1,2	1
17-jul-91	17	0,004	0,188	32	33	35	4,2	2
18-jul-91	15	0,094	0,495	38	37	35	4,6	2
19-jul-91	2	0	0,004	2	2	3	0,3	2
20-jul-91	2	0,02	0	4	7	6	0,7	2
21-jul-91	3	0	0,046	3	5	4	0,6	2
22-jul-91	5	0,008	0,023	5	7	4	0,8	2
23-jul-91	6	0,011	0,283	26	18	20	2,5	2
24-jul-91	7	0,08	0,022	13	15	14	1,8	2
25-jul-91	17	0,01	0,274	33	41	45	4,8	2
26-jul-91	22	0,015	0,16	43	44	43	5,5	2
27-jul-91	9	0,027	0,121	18	20	18	2,4	2
28-jul-91	12	0,02	0,243	40	28	23	3,7	2
30-jul-91	1	0,022	0,015	6	4	2	0,5	2
01-agosto-91	2	0	0,013	2	3	1	0,3	2
02-agosto-91	2	0	0,006	3	4	4	0,5	2
03-agosto-91	4	0	0,036	14	10	10	1,3	2
04-agosto-91	7	0,135	0,098	20	19	23	2,5	2
05-agosto-91	4	0,025	0,222	12	12	20	1,7	2
06-agosto-91	5	0	0,034	4	12	13	1,2	2
07-agosto-91	7	0,003	0,01	8	11	9	1,3	2
08-agosto-91	17	0,065	0,684	46	49	41	5,6	2
09-agosto-91	13	0,061	0,088	32	32	27	3,7	2
10-agosto-91	12	1,622	1,692	27	35	33	6,1	2
11-agosto-91	12	0,083	0,249	32	30	29	3,7	2
12-agosto-91	14	0,021	0,273	38	30	27	3,9	2
13-agosto-91	20	1,101	3,505	53	61	52	9,1	3
14-agosto-91	13	0,161	0,459	31	29	29	3,9	3
15-agosto-91	11	0,013	0,201	24	26	28	3,2	3
16-agosto-91	21	0,109	0,364	34	51	46	5,6	3
17-agosto-91	12	0,005	0,121	13	26	27	2,8	3
18-agosto-91	21	0,048	0,248	48	60	45	6,1	3
19-agosto-91	11	0	0,079	4	21	17	2,0	3
20-agosto-91	19	0,021	0,11	25	40	36	4,3	3
21-agosto-91	21	0,456	1,83	42	65	41	7,1	3
22-agosto-91	7	0,02	0,108	12	17	17	1,9	3
23-agosto-91	9	0	0,061	6	19	15	1,8	3
24-agosto-91	24	0,077	0,877	38	50	51	6,2	3
25-agosto-91	22	0,017	0,463	33	49	53	5,7	3
26-agosto-91	16	0,054	0,099	27	31	26	3,7	3
27-agosto-91	18	0,004	0,194	31	41	21	4,0	3
28-agosto-91	6	0	0,142	12	13	10	1,5	3
29-agosto-91	4	0,01	0,02	9	15	15	1,4	3
30-agosto-91	1	0	0,001	0	2	0	0,1	3
01-sept-91	6	0,005	0,064	5	16	13	1,4	3

02-sep-91	1	0,002	0	2	2	1	0,2	3
03-sep-91	2	0	0,004	5	5	3	0,5	3
04-sep-91	5	0	0,065	6	14	19	1,5	3
05-sep-91	1	0	0,01	0	1	0	0,1	3
06-sep-91	3	0,01	0,004	9	8	6	0,9	3
07-sep-91	7	0,01	0,03	12	13	11	1,6	3
08-sep-91	16	0,01	0,171	26	43	42	4,4	3
09-sep-91	11	0,033	0,117	22	26	29	3,1	3
10-sep-91	5	0	0,061	6	12	8	1,1	3

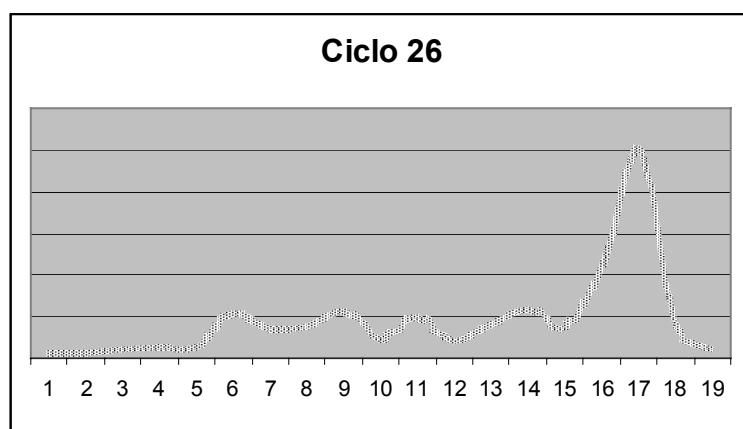
Tabla 53: Ciclo 15.



Día	NºInc	SupArb	SupRasa	Esp	Brig	MedMec	Oc.D.	Sector
15-feb-92	1	0	1,5	2	2	1	0,2	1
16-feb-92	1	0,2	0	1	2	1	0,2	1
18-feb-92	3	0	2,7	1	4	2	0,3	1
19-feb-92	9	0,1	10,9	4	13	5	1,1	1
20-feb-92	13	0,3	13,13	6	16	5	1,4	1
21-feb-92	19	6,1	27,52	11	29	9	2,9	1
22-feb-92	22	5,9	15,61	12	30	7	2,7	1
23-feb-92	42	2,5	66,7	14	61	18	5,1	2
24-feb-92	43	7,1	78,63	21	62	15	5,9	2
25-feb-92	35	3,5	82,31	30	54	17	5,7	2
26-feb-92	43	4,2	38,11	36	61	18	5,5	2
27-feb-92	60	21,9	39,9	29	84	27	7,9	3
28-feb-92	29	3,8	22,31	9	41	7	3,0	3
29-feb-92	34	7,7	66,1	13	57	15	5,1	3
01-mar-92	9	0,1	19,8	5	12	3	1,2	3
02-mar-92	6	1	7	3	13	4	0,9	3
03-mar-92	15	5,9	8,6	11	25	8	2,3	3
04-mar-92	36	9,1	38,1	20	60	17	5,1	3
05-mar-92	17	0,6	16	9	22	7	1,9	3
06-mar-92	35	1,9	56,3	36	47	26	5,5	3
07-mar-92	20	9,7	15,9	11	26	11	3,0	3
08-mar-92	21	2,7	8,7	4	26	10	2,0	3
09-mar-92	31	6,2	40,8	17	51	17	4,5	3
10-mar-92	24	4,7	50,8	18	36	17	4,1	3
11-mar-92	24	15	25,2	12	43	17	4,3	3

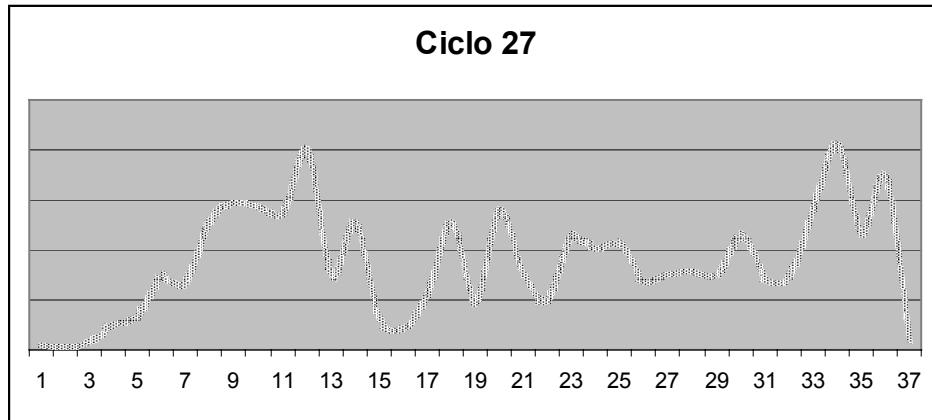
12-mar-92	15	6,7	12,1	14	30	14	2,8	3
13-mar-92	25	2,9	23,2	9	41	14	3,1	3
14-mar-92	20	2,6	25,45	19	38	13	3,2	3
15-mar-92	24	3,8	9,2	10	37	21	3,1	3
16-mar-92	23	8,6	23,75	24	50	24	4,6	3
17-mar-92	22	2,35	19,1	16	32	13	2,9	3
18-mar-92	23	4,6	14,5	13	34	14	3,0	3
19-mar-92	41	4,9	56,7	27	56	23	5,7	3
20-mar-92	44	15,95	43	45	78	43	8,2	3
21-mar-92	36	6,7	28,4	18	54	21	4,6	3
22-mar-92	47	3,8	61	39	72	30	6,9	3
23-mar-92	4	0	5,5	1	7	1	0,4	3

Tabla 54: Ciclo 24.



Día	NºInc	SupArb	SupRasa	Esp	Brig	MedMec	Oc.D.	Sector
23-ene-92	5	0	3,6	4	6	0	0,3	1
24-ene-92	3	0	4,6	4	5	2	0,3	1
25-ene-92	6	1,1	2,5	7	12	3	0,5	1
26-ene-92	8	0,5	8,8	3	13	4	0,5	1
27-ene-92	6	0,3	19	4	17	3	0,6	1
28-ene-92	28	10,1	52,5	28	39	8	2,2	1
29-ene-92	22	5	33,2	12	32	7	1,5	1
30-ene-92	23	2	26,45	18	33	6	1,5	1
31-ene-92	31	11,7	50,53	21	43	12	2,3	1
01-feb-92	16	5,9	12,4	7	17	5	0,9	1
02-feb-92	27	7,4	25,65	28	40	8	2,0	1
03-feb-92	11	1	34,63	8	15	4	0,8	1
04-feb-92	27	1,4	19,6	18	38	4	1,6	1
05-feb-92	27	1,5	125,8	27	45	8	2,4	1
06-feb-92	30	3	24,21	4	38	9	1,6	1
07-feb-92	53	25,2	128,61	36	87	27	4,5	2
08-feb-92	85	115,1	425,56	88	124	61	10,0	3
09-feb-92	13	14,8	66,5	11	25	11	1,6	3
11-feb-92	3	0,3	0,6	8	6	5	0,4	3

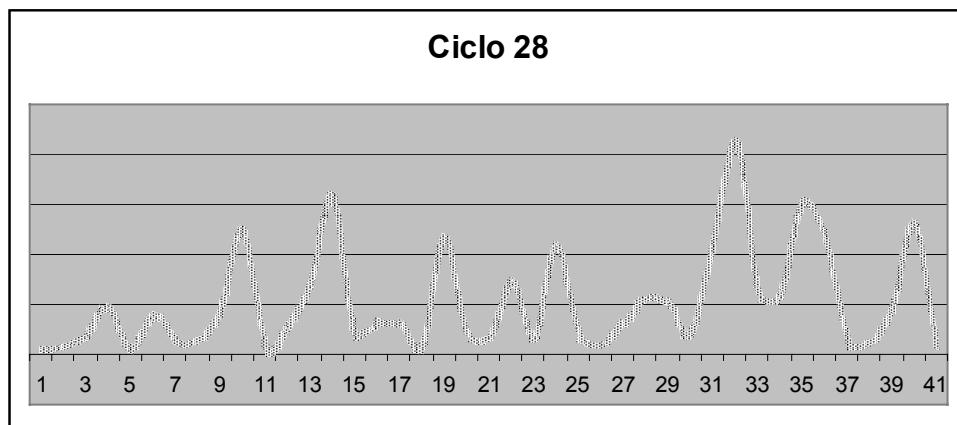
Tabla 55: Ciclo 26.



Día	NºInc	SupArb	SupRasa	Esp	Brig	MedMec	Oc.D.	Sector
15-feb-92	1	0	1,5	2	2	1	0,2	1
16-feb-92	1	0,2	0	1	2	1	0,2	1
18-feb-92	3	0	2,7	1	4	2	0,3	1
19-feb-92	9	0,1	10,9	4	13	5	1,1	1
20-feb-92	13	0,3	13,13	6	16	5	1,4	1
21-feb-92	19	6,1	27,52	11	29	9	2,9	1
22-feb-92	22	5,9	15,61	12	30	7	2,7	1
23-feb-92	42	2,5	66,7	14	61	18	5,1	2
24-feb-92	43	7,1	78,63	21	62	15	5,9	2
25-feb-92	35	3,5	82,31	30	54	17	5,7	2
26-feb-92	43	4,2	38,11	36	61	18	5,5	2
27-feb-92	60	21,9	39,9	29	84	27	7,9	3
28-feb-92	29	3,8	22,31	9	41	7	3,0	3
29-feb-92	34	7,7	66,1	13	57	15	5,1	3
01-mar-92	9	0,1	19,8	5	12	3	1,2	3
02-mar-92	6	1	7	3	13	4	0,9	3
03-mar-92	15	5,9	8,6	11	25	8	2,3	3
04-mar-92	36	9,1	38,1	20	60	17	5,1	3
05-mar-92	17	0,6	16	9	22	7	1,9	3
06-mar-92	35	1,9	56,3	36	47	26	5,5	3
07-mar-92	20	9,7	15,9	11	26	11	3,0	3
08-mar-92	21	2,7	8,7	4	26	10	2,0	3
09-mar-92	31	6,2	40,8	17	51	17	4,5	3
10-mar-92	24	4,7	50,8	18	36	17	4,1	3
11-mar-92	24	15	25,2	12	43	17	4,3	3
12-mar-92	15	6,7	12,1	14	30	14	2,8	3
13-mar-92	25	2,9	23,2	9	41	14	3,1	3
14-mar-92	20	2,6	25,45	19	38	13	3,2	3
15-mar-92	24	3,8	9,2	10	37	21	3,1	3
16-mar-92	23	8,6	23,75	24	50	24	4,6	3
17-mar-92	22	2,35	19,1	16	32	13	2,9	3
18-mar-92	23	4,6	14,5	13	34	14	3,0	3
19-mar-92	41	4,9	56,7	27	56	23	5,7	3
20-mar-92	44	15,95	43	45	78	43	8,2	3
21-mar-92	36	6,7	28,4	18	54	21	4,6	3
22-mar-92	47	3,8	61	39	72	30	6,9	3

23-mar-92	4	0	5,5	1	7	1	0,4	3
-----------	---	---	-----	---	---	---	------------	----------

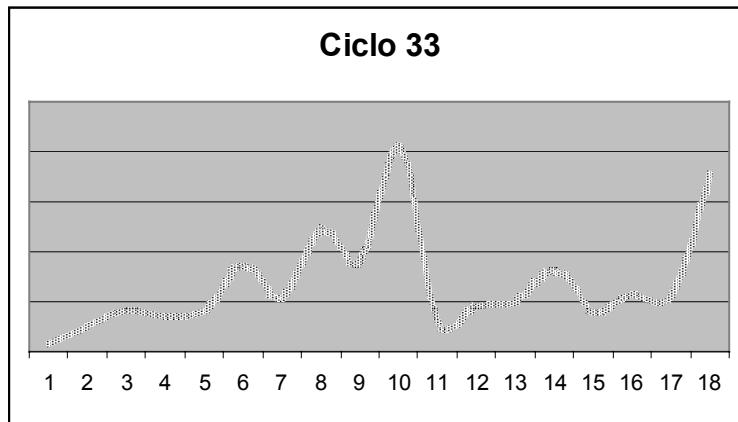
Tabla 56: Ciclo 27.



Día	NºInc	SupArb	SupRasa	Esp	Brig	MedMec	Oc.D.	Sector
10-abr-92	1	0,4	0,5	1	2	1	0,3	1
11-abr-92	1	0	1,5	1	2	1	0,3	1
12-abr-92	3	0	4,1	2	5	1	0,7	1
14-abr-92	6	3,3	9	9	11	3	1,9	1
15-abr-92	1	0	0,5	0	1	0	0,1	1
16-abr-92	6	2,2	4,8	8	10	3	1,5	1
17-abr-92	3	0	1,2	4	4	1	0,6	1
18-abr-92	3	0,5	0,5	4	4	1	0,6	1
19-abr-92	8	0,7	7,4	2	9	6	1,6	1
20-abr-92	14	1,3	37,6	21	24	13	5,0	2
21-abr-92	1	0,1	0,2	3	1	1	0,3	2
22-abr-92	5	0,2	2,5	7	6	4	1,1	2
23-abr-92	9	0,6	6,8	17	21	9	2,8	2
24-abr-92	16	19,1	29,5	20	34	18	6,3	2
25-abr-92	5	0,1	0,6	4	5	4	0,9	2
26-abr-92	5	0,5	2,7	9	7	5	1,3	2
27-abr-92	3	0,1	8,5	6	5	4	1,2	2
28-abr-92	1	0	2	0	3	1	0,3	2
29-abr-92	5	30	28,6	10	14	10	4,6	2
30-abr-92	5	0,1	1,2	3	7	2	0,9	2
02-may-92	2	2	0,1	7	4	2	0,7	2
03-may-92	6	10,1	12,2	13	14	8	2,9	2
04-may-92	3	2	0,7	2	4	1	0,6	2
05-may-92	11	2,7	20,8	23	24	17	4,3	2
06-may-92	4	1	3,6	2	7	2	0,9	2
07-may-92	2	0	0,3	3	2	2	0,4	2
08-may-92	5	0,1	2,7	6	8	5	1,2	2
09-may-92	7	1,3	6	14	13	9	2,2	2
10-may-92	7	0,3	11,6	5	13	9	2,1	2
11-may-92	3	1,6	0,7	3	3	5	0,8	2
12-may-92	15	2,2	7,9	17	29	20	4,2	2
13-may-92	23	22,1	12,9	50	46	34	8,5	3
14-may-92	5	4,1	9,9	14	16	9	2,5	3

15-may-92	8	1,2	1	8	19	14	2,3	3
16-may-92	17	2,2	32,9	19	34	27	6,0	3
17-may-92	22	2	6,6	17	35	22	4,9	3
18-may-92	1	0	2,5	3	2	3	0,5	3
19-may-92	2	0	0,5	3	3	2	0,5	3
20-may-92	6	0,5	6,4	7	10	8	1,7	3
21-may-92	17	24,5	4,4	14	31	16	5,2	3
23-may-92	1	0,1	0	3	2	2	0,3	3

Tabla 57: Ciclo 28.



Día	NºInc	SupArb	SupRasa	Esp	Brig	MedMec	Oc.D.	Sector
21-jul-92	1	0	0,1	2	2	1	0,3	1
22-jul-92	3	0,45	0,25	6	6	5	1,0	1
23-jul-92	3	0,8	1,8	8	8	11	1,7	1
24-jul-92	5	0	1,22	6	7	9	1,5	1
25-jul-92	5	0,4	1,18	12	6	8	1,7	1
26-jul-92	8	1,15	3,62	16	19	21	3,5	1
27-jul-92	2	0	13,8	7	8	8	2,2	1
28-jul-92	9	1,95	23,39	12	20	20	4,9	2
29-jul-92	10	2,35	6,71	17	18	13	3,6	2
30-jul-92	23	15,4	4,47	37	44	23	8,2	3
31-jul-92	4	0,75	0,08	3	6	7	1,1	3
01-agosto-92	8	1,25	1,15	4	12	8	1,8	3
02-agosto-92	7	0,5	0,72	13	11	7	2,0	3
03-agosto-92	9	0,3	5,21	21	16	13	3,3	3
04-agosto-92	5	0,16	2,62	4	10	9	1,6	3
05-agosto-92	6	0,27	3,41	7	20	10	2,3	3
06-agosto-92	6	0	5,39	9	11	12	2,3	3
07-agosto-92	19	5,07	7,78	29	42	32	7,0	3

Tabla 58: Ciclo 33.

Analizando estos ciclos y aplicando funciones que los resuman, se llega a la definición del prototipo “**Siniestralidad Altamente Progresiva**”, que es el que refleja la tabla 59:

	Sector 1	Sector 2	Sector 3
Media de días:	8	13	15
Media de incendios/día:	9	12	19
Mínimo de inc/día:	1	1	1
Máximo de inc/día:	31	53	85
Nº especialistas/día:	8	16	17
Nº de brigadas/día:	13	21	33
Nº de medios/día:	5	13	18

Tabla 59: Prototipo “Siniestralidad Altamente Progresiva”.

1.4. Representación borrosa de los prototipos.

Los prototipos han sido representados como “números borrosos”, los cuales permitirán obtener un grado de pertenencia, un grado de compatibilidad si es que existe, de una situación real con cada uno de los prototipos anteriormente calculados. En bien de la sencillez del modelo han sido representados como números borrosos triangulares, isósceles e iguales. Por tanto, para representar los prototipos (números borrosos triangulares) sólo se necesita conocer su centro (“Centro de prototipo”) y el tamaño de la base del triángulo.

El problema de caracterizar la representación borrosa de los prototipos surge en el proceso de adquisición del conocimiento experto; más concretamente, del hecho de que el experto no es capaz de dar reglas que asociasen unos valores de los factores con el prototipo, pero si es capaz de mostrar ejemplos o casos que caen bajo un prototipo determinado. Dicho de otro modo, el experto no realizaba un proceso deductivo, pero sí un proceso de exemplificación.

Como ya se ha explicado anteriormente, la evolución de la siniestralidad viene dada por ciclos que se inician a partir de cada periodo de no ocurrencia de incendios (normalmente motivada por lluvias). Los factores que determinan *a priori* la progresión de la siniestralidad desde un periodo de lluvias al siguiente se pueden concentrar en tres. El usuario introducirá sus valores al principio de cada ciclo:

A. Sequedad estacional.

La *sequedad estacional* es la propia de la época del año; entendiendo que puede haber un Invierno seco o un Verano lluvioso. {Alta/0.95, Media/0.45, Baja/0.05}.

B. *Intensidad de la última lluvia.*

En nuestro caso, los expertos consideran que la influencia de los tipos de intensidad vendría dada por {Llovizna/0.9, Tormenta/0.5, Borrasca/0.1}.

C. *Influencia social.*

Representa la influencia del factor humano en la evolución de la siniestralidad (conflictividad de la zona, rencillas entre vecinos, etc) en un ciclo (posteriormente se verá que también puede haber influencia social puntual, por ejemplo un día señalado). El usuario determina el tipo de influencia por {Mucha/0.85, Media/0.45, Escasa/0.05}.

Como primera aproximación se entiende que algún tipo de media (pensada como función de agregación de información) de los casos dados por el experto podría servir para determinar el centro de cada prototipo.

El proceso para determinar los centros de los prototipos es el siguiente:

- I. Se le presenta al experto una batería de casos elegidos al azar y se le pide que señale el prototipo bajo el que cae cada caso. Véase un ejemplo en la tabla 60:

Sequedad Estacional	Intensidad Lluvia	Influencia Social	Siniestralidad Progresiva	Media
Alta	Llovizna	Mucha	Elevada	0.900
Alta	Borrasca	Mucha	Media	0.633
Media	Tormenta	Media	Media	0.467
Media	Llovizna	Mucha	Elevada	0.733
Baja	Borrasca	Media	Escasa	0.200
Baja	Llovizna	Mucha	Media	0.600
Alta	Tormenta	Escasa	Media	0.500
Media	Borrasca	Escasa	Escasa	0.200
Baja	Tormenta	Baja	Escasa	0.200

Tabla 60.

Las tres primeras columnas señalan la información que se le da al experto (los casos); la cuarta es rellenada por el experto (los prototipos). Con el fin de respetar el lenguaje del experto sólo se manejan etiquetas lingüísticas en la entrevista, sin mostrarle las medias (quinta columna).

- II. Si algún prototipo no tuviese casos o no los suficientes se realiza el proceso inverso, es decir, preguntar al experto por ejemplares de dichos prototipos. Ejemplo:

Como se observó que había pocas ocurrencias de los prototipos “Progresión elevada” y “Progresión escasa”, se le pidió al experto que enunciase casos de estos prototipos (tabla 61):

Sequedad Estacional	Intensidad Lluvia	Influencia Social	Siniestralidad Progresiva	Media
Alta	Tormenta	Mucha	Elevada	0.767
Alta	Llovizna	Media	Elevada	0.767
Media	Borrasca	Media	Escasa	0.333

Tabla 61.

- III. Por tanto, los centros de prototipos son las medias de los casos de cada prototipo. En este ejemplo:

Centro de “Progresión elevada”: 0.791

Centro de “Progresión media”: 0.550

Centro de “Progresión escasa”: 0.233

- IV. El último paso es la representación de los números borrosos triangulares, a partir de dos factores:

A. El centro de cada prototipo.

B. La distancia del cero al centro del primer prototipo se usa como distancia estándar para construir la base de los tres triángulos (fig. 35).

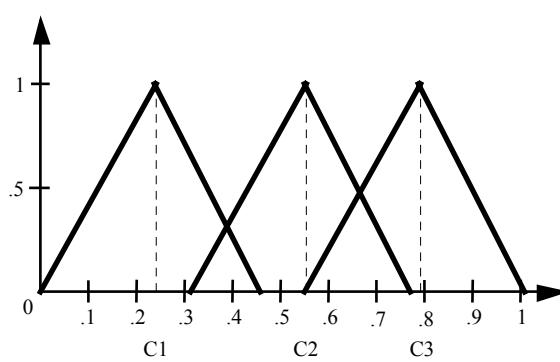


Figura 35.

1.4.1. Generación y representación de números borrosos triangulares.

Se presenta un interfaz donde se pueden consultar los centros de los números borrosos existentes y el nombre del prototipo al que representan. Los valores de los centros pueden modificarse lo cual obliga a un ajuste del segundo parámetro que define a los números borrosos, o sea, la amplitud de la base de los triángulos. Todos estos datos residen en una tabla relacional denominada Prototipos, a la cual se accede mediante clases derivadas de la MFC CRecordSet.

Nombre del Campo	Tipo de Datos	Descripción
CODIGO	autonumérico	Clave principal. Identificador en BD.
NOMBRE	texto	Nombre del prototipo representado.
CENTRO	real simple	Centro del triángulo.
EXTREMO_A	real simple	Extremo inferior con valor Y=0.
EXTREMO_B	real simple	Extremo superior con valor Y=0.

Tabla 62: Definición de números borrosos 1.

Por defecto la distancia del cero al primer prototipo se usa como distancia estándar para construir la base de los tres triángulos, pero pueden presentarse ocasiones en que este comportamiento habitual tenga que ser modificado:

Sean C1, C2 y C3 los centros de los tres números borrosos ordenados por su proximidad al cero (el nº borroso C1 es el más próximo); si se tiene en cuenta que están normalizados, son triángulos isósceles puros entre 0 y 1 y no trapecios, entonces si C3 + C1 no llega a ser igual o mayor que 1, la distancia a C1 no sirve como valor para construir la base de los triángulos si no que se tendría que escoger un nuevo valor que sería 1 – C3, que sí cumpliría la condición de cobertura total del 0 a 1 por parte de los números borrosos.

Cada vez que se produzca un cambio en los centros se recalculará la base y se dibujarán de nuevo en el panel que a tal efecto existe en el interfaz de los triángulos representativos.

- | |
|---|
| 1. Extraer candidatos para amplitud |
| 1.1 Tomar centro del primer número borroso como extremo inferior |
| 1.2 Tomar centro del último número borroso como extremo superior |
| 2. Comprobamos candidatos |
| 2.1 Si centro primer número válido entonces amplitud = centro primer nº borroso |
| 2.2 Si no amplitud = 1 – centro último nº borroso |
| 3. Actualizar amplitudes de triángulos. |

Algoritmo 10: Pseudocódigo de actualización de amplitudes.

```

m_SProt.MoveFirst();
/* se toma el centro C1 (el primero ) y C3 el último de la base de datos */
double extrinf = m_SProt.m_Centro;
m_SProt.MoveLast();
double extrsup = m_SProt.m_Centro;

/* se calcula la amplitud de los números borrosos triangulares
 * primero probando con el primer centro y si se queda corto
 * se prueba con el 1 - ultimo centro
 */
// se comprueba si vale el primer centro
if(extrsуп + extrinf > UNO ) {
    // si es así se actualizan los valores de amplitud de todos los prototipos
    m_SProt.MoveFirst();
    while (!m_SProt.IsEOF()) {
        Actualizar(&m_SProt,extrinf);
        m_SProt.MoveNext();
    }
}

else { // si no vale el primer centro será 1 - último centro
    double ampl = UNO - extrsup;
    // se actualizan los valores de amplitud de todos los prototipos
    while (!m_SProt.IsEOF()) {
        Actualizar(&m_SProt,ampl);
        m_SProt.MoveNext();
    }
}

// se manda redibujar el panel de triángulos
Invalidate();

```

Algoritmo 11: Adaptación de amplitudes de números borrosos.

La función Actualizar modifica todos las definiciones de prototipos que existen en la tabla en función de la amplitud que se le pase como parámetro.

A partir de los datos que residen en la tabla de prototipos se pueden representar gráficamente éstos en un panel que existe en el interfaz para tal propósito. Para ello se utiliza la función OnPaint() que poseen todas las aplicaciones Windows y se le añade el código que contiene a continuación, que se ejecutará cada vez que necesite ser modificada esta representación gráfica (invocando a Invalidate).

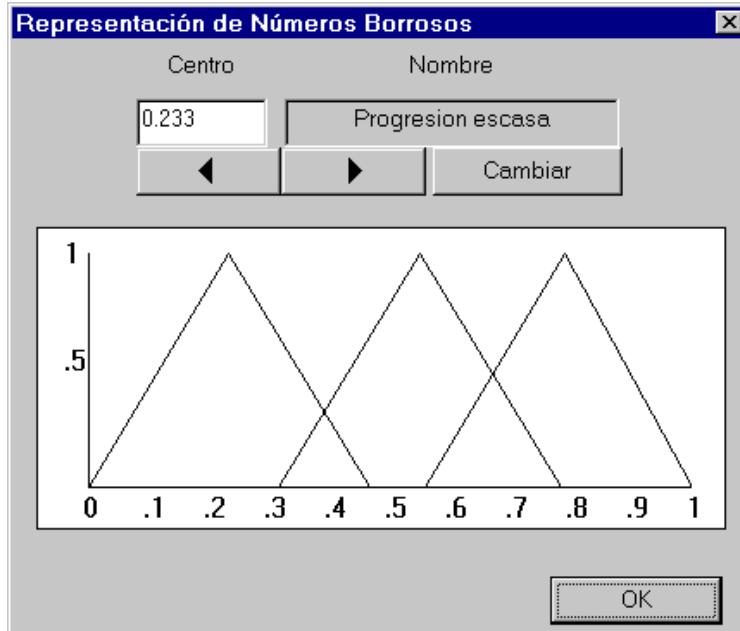


Figura 36: Interfaz de la aplicación.

1.5. Otro punto de vista para la determinación de los prototipos.

Como segunda aproximación, parece natural el uso de algoritmos de aprendizaje por inducción a partir de ejemplos para clasificar bajo qué clase cae un nuevo caso. Para ello se utilizan los algoritmos ID3 y CART, explicados con detalle en el Apéndice 1, de los que ahora conviene resaltar algunos aspectos relevantes para su implementación y uso en el caso real.

¿Cómo se representa un ejemplo?

Un ejemplo se representa como una clase (en sentido de programación, *objeto*) que contiene una lista de atributos con sus valores correspondientes y un valor de clase (en sentido de *clasificación*, clase a la que pertenece):

- Propiedades:

Clase: clase del ejemplo.

Lista de Atributos del ejemplo.

- Métodos:

Constructores.

Generadores (introducción de valores de atributos y clases).

Métodos de *consulta* de valores.

Operadores sobrecargados (=).

¿Cómo se representa un separador?

Un separador se representa como una clase que contiene dos valores: Un número de atributo y el valor de éste, lo cual conjuntamente servirá para poder dividir un conjunto según un criterio.

- Propiedades:

Índice del atributo que funciona como separador.

Valor concreto del atributo para el separador.

- Métodos:

Constructores y destructores.

Operadores sobrecargados (igualdad y comparación).

¿Cómo se representa un árbol de decisión (regla)?

Una regla es el resultado que devuelven los algoritmos de clasificación: La expresión del árbol de decisión. Contiene los datos necesarios para representar los resultados de CART e ID3 para posibilitar la posterior evaluación de ejemplos (A partir de los valores de los atributos inducir la clase). Es parecida a la clase separador:

- Propiedades:

Índice del atributo que funciona como separador.

Valor concreto del atributo para el separador.

Lista de Reglas Hijas.

- Métodos:

Constructores y destructores.

Operadores sobrecargados.

Métodos de *consulta*.

¿Cómo se representa el Conjunto de Ejemplos?

Un conjunto de ejemplos se representa como una lista de ejemplos, cuyos atributos tienen unos nombres (*campos*) a los cuales se les van aplicando una serie de separadores.

- Complementos:

Lista de Ejemplos.

Lista de Campos (valores de cadenas).

Lista de Separadores.

El proceso de construcción/carga.

Los ejemplos se presentan en una tabla relacional en formato MS-Access '97 por lo que son fácilmente accesibles mediante clases derivadas de la MFC CRecordSet utilizando ODBC. La tabla de ejemplos se acompaña con una tabla de campos con los nombres de los atributos que caracterizan a cada ejemplo. Así se podrá contar para la presentación de la información con las etiquetas y los valores de cada separador.

Para pasar de los datos de la tabla a un objeto de la clase CCjtoEjemplos, se hace que el constructor del objeto reciba como parámetros dos punteros a clases CrecordSet, o sea dos punteros a tablas relacionales, uno para los ejemplos y otro para los campos. De esta forma con el constructor se recorren las dos tablas para almacenar en memoria toda la información que ofrezca la Base de Datos y poder trabajar con ella sin tener que acceder más veces, posibilitando así la obtención de la solución de una forma rápida con un mínimo coste en almacenamiento.

1.Hasta final Tabla Ejemplos

- 1.1 Crear un nuevo Ejemplo
- 1.2 Para Todo Campo que sea Atributo
 - 1.2.1 Tomar el valor de Atributo
 - 1.2.2 Almacenarlo en el ejemplo
- 1.3 Tomar la clase del ejemplo y almacenarla
- 1.4 Insertar Ejemplo en el Conjunto de Ejemplos
- 1.5 Avanzar al siguiente registro de la tabla

2. Hasta el final de la Tabla Campos

- 2.1 Almacenar el valor del Campo
- 2.2 Avanzar al siguiente registro

3. Establecer nº de separadores utilizados a cero.

Algoritmo 11: Pseudocódigo del constructor de carga de la BD.

El constructor por defecto es diferente ya que la carga de la BD sólo se debe hacer al comenzar el algoritmo. Al crear subconjuntos a partir de un separador, primero se construye un conjunto de ejemplos vacío para ir insertándole los ejemplos que le correspondan, según la aplicación del separador.

Los métodos de consulta.

Para manipular un conjunto de ejemplos mediante algoritmos de clasificación se deben implementar unos métodos generales de consulta para obtener datos específicos sobre él, como el número de clases diferentes del conjunto, el número de casos por clase ..., datos que van a ser necesarios en la aplicación de los dos algoritmos.

```
/* MÉTODOS DE CONSULTA */

int EsVacio() { return m_NumEjemplos == 0; }

// método de consulta que determina el nº de ejemplos que contiene el cjto
int TamanyoCjto() { return m_NumEjemplos; }

// devuelve el número de clases del conjunto
int NumClases();
// devuelve el número de clases del conjunto y además
// un vector con los valores de esas clases
int TomarClases(CString VectClases[MAXCLASES]);

// devuelve el número de apariciones de una clase
// en el conjunto de ejemplos
int CasosClase(CString clase);

// devuelve el número de valores diferentes que posee
// un atributo en un conjunto de ejemplos y un vector
// con esos valores
```

```

int NumValoresDif(int nAtr, CString Valores[MAXVALORES]);

/* métodos de consulta sobre ejemplos individuales */

// devuelve la clase de un ejemplo determinado
CString ClaseEjemplo(int dx) {return m_Ej[dx].ValorClase();}

// devuelve el valor de un atributo dentro de un ejemplo
CString ValorAtrEjemplo(int dx, int nAtr) {
    Return m_Ej[dx].ValorAtributoDX(nAtr);
}

```

Algoritmo 12: Métodos de consulta.

Otros métodos auxiliares importantes son los métodos de división en subconjuntos:

- *ObtenerSubCjto()*: Obtiene un subconjunto insertándole los ejemplos del conjunto inicial que cumplan las condiciones de un separador (un atributo tiene un valor concreto). Para que se pueda aplicar el algoritmo (ID3 en este caso) sobre él, se le debe indicar cual ha sido el separador utilizado, lo cual lo realiza el método *AplicarCampos()*.
- *Dividir()*: Divide el conjunto de ejemplos en dos subconjuntos a partir de los datos de un separador, uno de ellos cumple los requisitos del separador y el otro no. Para que no se vuelvan a aplicar en estos subconjuntos los separadores ya empleados se utilizar el método *PropagarSeparadores()*.

1.5.1. ID3.

A: ID3 (devuelve una regla)

- Calcular el número de clases del conjunto de ejemplos
- Si el conjunto de ejemplos tiene más de una clase
 - Calcular el atributo mejor separador
(Llamada a *CalcularMejorSeparador*);
 - Generar los hijos a partir de los diferentes valores que pueda tomar ese atributo
(Diferentes llamadas a *ObtenerSubCjto*)

B: CALCULAR MEJOR SEPARADOR

(devuelve el atributo del mejor separador)

- Calcular la entropía raíz
Entropia_raiz = CalcularEntropia();
- Obtener tamaño del conjunto: Tamaño total
- Para cada atributo que no se haya utilizado como separador
 - Se obtienen los distintos valores que tiene ese atributo

- B.3.2. Se divide en subconjuntos según los valores obtenidos
 B.3.3. Se calcula la entropía de cada subconjunto
 B.3.4. Se calcula la entropía relacionada con el separador

$$(E = \text{Sumatorio}((\text{Tamaño del Subconjunto} / \text{tamaño total}) * \text{Entropía Subconjunto}))$$

 B.3.5. Calcular el decremento de entropía

$$\text{Dec_}E = \text{Entropía_raiz} - E$$

 B.3.6. Comprobar si es el mayor decremento de entropía hasta el momento y si es así se almacena
 B.4 Devolver como resultado el índice del atributo que ha resultado mejor separador.

C: CALCULAR ENTROPIA

(devuelve la entropía del conjunto de ejemplos)

- C.1 Calcular el número de casos por clase (método CasosClase(clase))
 C.2 Inicializar Entropía a 0
 C.3 Para cada Clase

$$\text{C.3.1 Entropía} = \text{Entropía} - ((\text{casos_clase} / \text{tam_cjto}) * \log_2(\text{casos_clase}/\text{tam_cjto}))$$

 C.4 Devolver Entropía

Algoritmo 13: Pseudocódigo de ID3.

```

CRegla * CcjtoEjemplos::id3Recur()

{
/* B.1 Calcular el número de clases del conjunto ejemplo */
CString clases[MAXCLASES];
int nNumClases = TomarClases(clases);

CRegla* R = new Cregla();

if (nNumClases > 1) {
    // atributo mejor separador
    int nAtributo;

    /* Llamar Calcular_mejor_separador, que
       Devuelve el mejor separador (atributo+umbral) */
    Natributo = CalcularMejorSeparador();

    Cstring Valores[MAXVALORES];
    // numero de valores posibles del mejor separador
    int      nNumHijos = NumValoresDif(nAtributo,Valores);

    /* se guarda la solución */
    R->separador = Campos[nAtributo];
    R->nNumHijos = nNumHijos;
    /**
    CcjtoEjemplos SubCjto[MAXSUBCONJ];

    int ii;
    for (ii = 0; ii < nNumHijos; ii++) {
        ObtenerSubCjto(&SubCjto[ii],nAtributo,Valores[ii]);
        SubCjto[ii].Campos[nAtributo] = "";
        R->Hijos[ii] = SubCjto[ii].id3Recur();
        CString aux;
        Aux.Format("%s---%s",Valores[ii],R->Hijos[ii]->separador);
    }
}
}

```

```

        R->Hijos[ii]->separador = aux;

    }

    return R;

}

else {/* Si no*/
    /* Grabar en el fichero de salida la clase de los elementos
       del conjunto ejemplo */
    if (nNumClases == 1) R->separador = "@@@" + clases[0] + "@@@";
    else R->separador = "";
    R->nNumHijos = 0;
    Return R;

}

return R;
}

```

Int CCjtoEjemplos::CalcularMejorSeparador()

```

{
    /* VARIABLES LOCALES (luego se colocaran en su lugar) */
    // entropía raíz del conjunto ejemplo
    double E_raiz;
    double MaxDecr = 0;
    int AtrMejor;

    /* Llamar Calcular_entropia() para obtener la entropía raíz */
    E_raiz = CalcularEntropia();
    // calcular el tamaño del conjunto
    int tam = TamanyoCjto();
    //

    /* Para cada atributo */
    int i;
    for (i = 0; i < (int) m_NumCampos; i++){
        if (Campos[i] == "") continue;
        /*. Para cada caso por clase */
        /* el número de casos por clase */
        Cstring Valores[MAXVALORES];
        Int nNumVDif = NumValoresDif(i, Valores);

        CcjtoEjemplos SubCjto[MAXSUBCONJ];

        /* se divide en subconjuntos */
        int ii;
        for (ii = 0; ii < nNumVDif; ii++) {
            ObtenerSubCjto(&SubCjto[ii], i, Valores[ii]);
        }
        /* se calcula la entropía de los subconjuntos */
        double ValEntropia[MAXSUBCONJ];

        for (ii = 0; ii < nNumVDif; ii++)

```

```

ValEntropia[ii] = SubCjto[ii].CalcularEntropia();

Double dE = 0;
/* se calcula entropía general */
for (ii= 0; ii < nNumVDif; ii++)
    dE+=((double) SubCjto[ii].TamanyoCjto() / (double) tam)
        * ValEntropia[ii];
/* Dec_E = Entropía_raiz - E */
double Dec_E = (double) E_raiz - dE;

/* Guardar separador si Dec_E es mayor que los anteriores */
if (Dec_E >= MaxDecr){
    AtrMejor = i;      MaxDecr = Dec_E;
}
} // fin para cada atributo
/* C.3 Devolver el índice del mejor separador */
return AtrMejor;
}

```

Double CcjtoEjemplos::CalcularEntropia()

```

{
    int nCasosXClase[MAXCLASES];
    // tamaño del conjunto
    int nTamanyo=TamanyoCjto();
    double dEntropia;

    /* Calcular el número de casos que hay por clase */
    // se calcula el número de clases y se toman sus valores
    Cstring strClases[MAXCLASES];
    Int nNumClases = TomarClases(strClases);
    Int i;
    For (i =0; i < nNumClases; i++)
        nCasosXClase[i] = CasosClase(strClases[i]);

    /* Entropía = 0 */
    dEntropia = 0.00;

    /* D.3 Para cada clase */
    for (i = 0; i < nNumClases; i++)
        if (nCasosXClase[i] != 0 )
            /* Entropía = Entropía - ((casos_clase/tam_cjto)*
               log2(casos_clase/tam_cjto) */
            dEntropia = dEntropia - (((double) nCasosXClase[i] / (double)
                nTamanyo )*(log((double) nCasosXClase[i] / (double) nTamanyo)) /
                log((double)2)) ;

    /* Devolver entropía */
    return dEntropia;
}

```

Algoritmo 14: Implementación de los métodos del algoritmo ID3.

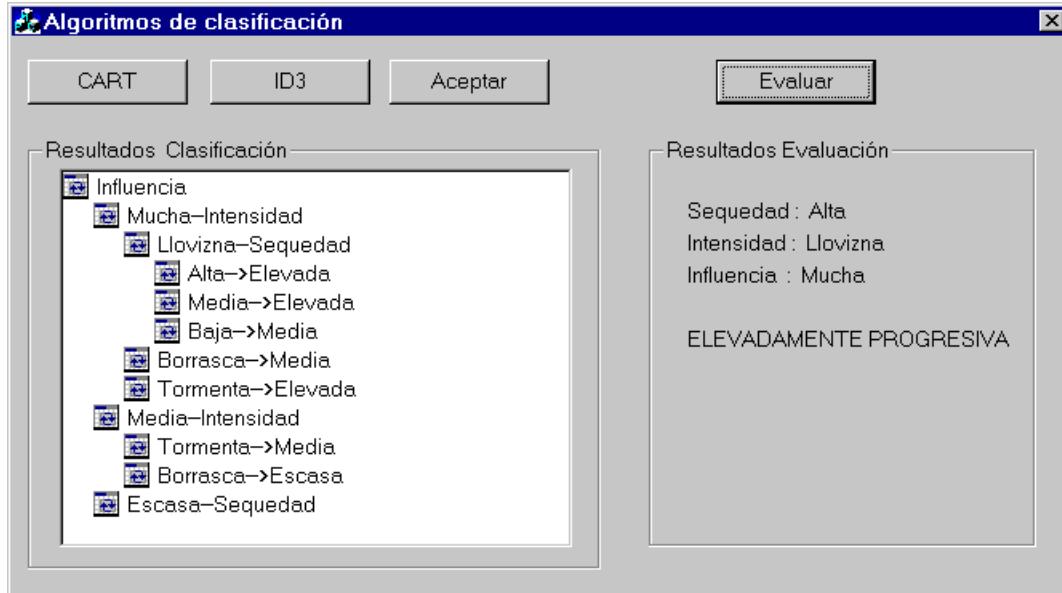


Figura 37: Interfaz y ejemplo del algoritmo ID3.

1.5.2. CART.

A. CART (versión sin poda) (devuelve una regla)

A.1. Se calcula la impureza del nodo (Conjunto de ejemplos)

A.2. Si es un conjunto impuro (impureza > 0)

A.2.1 Se Calcula el mejor separador

Invocación a CalcularMejorSeparador

A.2.2 Se almacena separador en la regla y se apunta como separador utilizado

A.2.3 Se parte en dos el conjunto según el mejor separador

(método Dividir)

A.2.4 Aplicar CART a cada Subconjunto

A.3. Si no es impuro entonces (impureza = 0)

A.3.1 Es un conjunto de una clase : Se almacena la clase resultante

A.3.2 Es un conjunto vacío: se indica en la regla

A.4 Se retornan reglas

A. CART (versión con poda, la única variación se sitúa en el punto A.2.4)

A.2.4 Antes de llamar recursivamente a CART para cada SubConjunto

A.2.4.1 Se calcula la tasa de ramificación del conjunto actual

A.2.4.2 Se calcula la tasa de ramificación de los SubConjuntos

A.2.4.3 Si la tasa de ramificación del Subconjunto 1 es menor que la del conjunto entonces se aplica CART sobre él, si no se detiene la expansión

A.2.4.4 Si la tasa de ramificación del Subconjunto 2 es menor que la del conjunto entonces se aplica CART sobre él, si no se detiene la expansión

Algoritmo 15: Pseudocódigo de CART.

B. CalcularMejorSeparador (versión CART)

(Devuelve como resultado un separador compuesto por un número de atributo y un valor concreto de ese atributo)

B.0 Inicialización de los valores de almacenamiento del mínimo de impureza**B.1 Obtener los separadores que aún no han sido aplicados****B.2 Para cada separador obtenido**

- B.2.1 Dividir el subconjunto según el separador (*método dividir*)
- B.2.2. Para cada subconjunto (que son 2) se calcula su impureza
- B.2.3 Obtener la impureza del separador
Mínimo(impurezaSubCjto1, impurezaSubCjto2).
- B.2.4 Si la impureza es menor que la que hasta ahora era menor
- B.2.5. Almacenar el separador como mejor provisional

B.3 Devolver mejor separador**C. IMPUREZA** (según el criterio de GINI)- *Cálculos iniciales*

- C.1. Tamaño del Conjunto
- C.2. Calcular el número de casos que hay por clase

- *Impureza por Criterio de Gini*

- C.3. Inicializar Impureza a 0

C.4. Impureza = Sumar

- C.4.1 Prob (nodo de clase i) * Prob (nodo de clase j)
Siempre que i no sea igual a j.

C.5. Regresar Impureza

Algoritmo 16: Pseudocódigo de los métodos de CART.

```
CRegla *CCjtoEjemplos::cart()

    CRegla* R= new CRegla();
    /* 1. Calcular la impureza el nodo (Conjunto de ejemplos */
    double d = Impureza();
    /* 2. Si es un conjunto Impuro */
    if(d != 0){
        CString ValorSep;
        int nAtributo;
        /* 2.1 se calcula el mejor separador */
        CalcularMejorSeparador(&nAtributo,&ValorSep);
        // 2.2 almacenar separador en la regla
        CString str;
        str.Format("%s->%s", Campos[nAtributo], ValorSep);
        R->separador = str;
        R->nNumHijos = 0; // 2
        // y en el conjunto de separadores utilizados
        m_Sep[m_NumSeparadores].m_NumAtributo = nAtributo;
        m_Sep[m_NumSeparadores++].m_Valor = ValorSep;
```

```

/* 2.3 Partir en dos el conjunto según el mejor separador */
CCjtoEjemplos *pCEIzdo, *pCEDcho;
pCEIzdo = new CCjtoEjemplos();
pCEDcho = new CCjtoEjemplos();
Dividir(pCEIzdo,pCEDcho,nAtributo,ValorSep);
/* 2.4 Llamar recursivamente a CART para cada
* subconjunto
*/
R->Hijos[R->nNumHijos++] = pCEIzdo->cart();
R->Hijos[R->nNumHijos++] = pCEDcho->cart();
/* 2.4 VERSIÓN CON PODA
// no ramificaremos si la tasa de branching es mayor
double Rn = TasadeRamificacion();
double RIzdo = pCEIzdo->TasadeRamificacion();
double RDcho = pCEDcho->TasadeRamificacion();
if (RIzdo < Rn) R->Hijos[R->nNumHijos++] = pCEIzdo->cart();
if (RDcho < Rn) R->Hijos[R->nNumHijos++] = pCEDcho->cart();
*/
}
else {
    /* Si es un conjunto puro entonces se almacena la
    * clase resultante y si es un conjunto vacío un ? */
    if (TamanyoCjto() == 0) R->separador = "?";
    else {
        CString clases[MAXCLASES];
        int nNumClases = TomarClases(clases);
        R->separador = clases[0] + "mente Progresiva";
    }
    R->nNumHijos = 0;
}
return R;
}

```

```

Void CcjtoEjemplos::CalcularMejorSeparador(int * pnAtr, CString * valor)
{
    /* 0. Inicialización de variables locales */
    double MinImp = 1000;
    *pnAtr = -1;
    /* obtener los separadores que se pueden aplicar */
    Cseparador separadores[MAXCAMPOS*MAXVALORES];
    int nSeparadores = ObtenerSeparadores(separadores);
    int i;
    /* Para cada separador obtenido */
    for ( i = 0; i < nSeparadores; i++) {
        CcjtoEjemplos SubCjto1, SubCjto2;
        /* Dividir en subconjuntos según el separador */
        Dividir(&SubCjto1,&SubCjto2,
                separadores[i].m_NumAtributo,separadores[i].m_Valor);

        /* determinar la impureza de los subconjuntos */
        double ImpIzq = SubCjto1.Impureza();
        double ImpDer = SubCjto2.Impureza();
        /* como impureza del separador se elige la mínima de los
        * subconjuntos
        */
    }
}

```

```

        double Imp = (ImpIzq <= ImpDer)? ImpIzq.ImpDer;
        /* Memorizar el separador que lleva a una impureza mínima */
        if (Imp <= MinImp){
            *pnAtr = separadores[i].m_NumAtributo;
            *valor = separadores[i].m_Valor;
            MinImp = Imp;
        }
    }

Double CCjtoEjemplos::Impureza() {
    int nCasosXClase[MAXCLASES];
    // tamaño del conjunto
    int nTamanyo=TamanyoCjto();

    /* Calcular el número de casos que hay por clase */
    CString strClases[MAXCLASES];
    int nNumClases = TomarClases(strClases);
    int i;
    for (i =0; i < nNumClases; i++)
        nCasosXClase[i] = CasosClase(strClases[i]);

    /* D.2 Impureza = 0 */
    double dImpureza = 0.00;

    int ii,jj;

    /* calculo de la impureza por el CRITERIO DE GINI */
    for ( ii = 0; ii < nNumClases; ii++)
        for (jj = 0; jj < nNumClases; jj++)
            if (ii != jj) dImpureza += ((double) nCasosXClase[ii] / (double)
                m_NumEjemplos) * ((double) nCasosXClase[jj] / (double)
                m_NumEjemplos);

    return dImpureza;
}

```

Algoritmo 17: Implementación de CART.

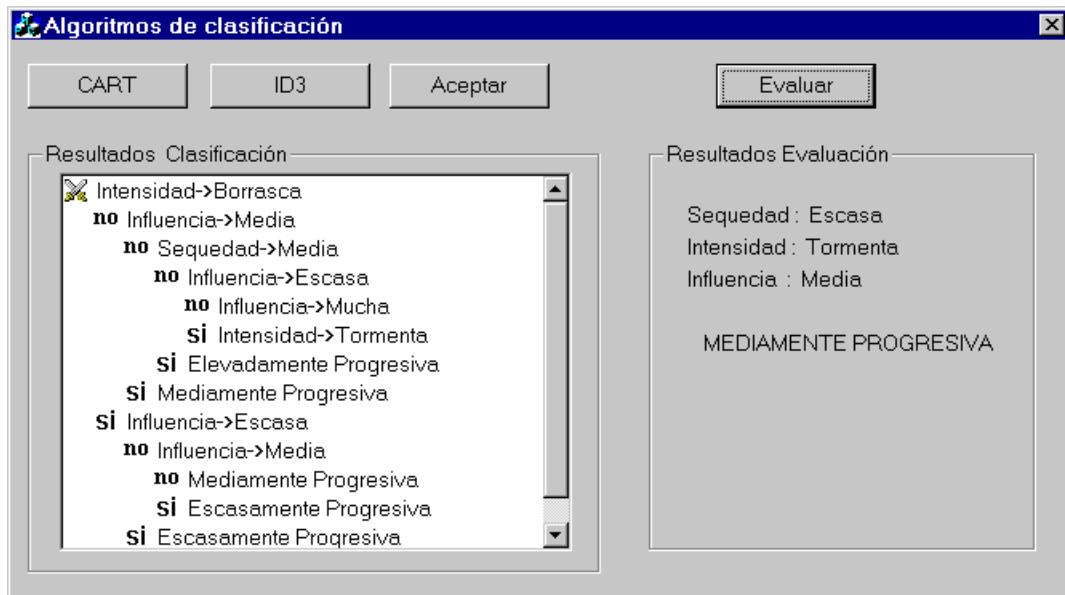


Figura 38: Interfaz y ejemplo de CART sin poda.

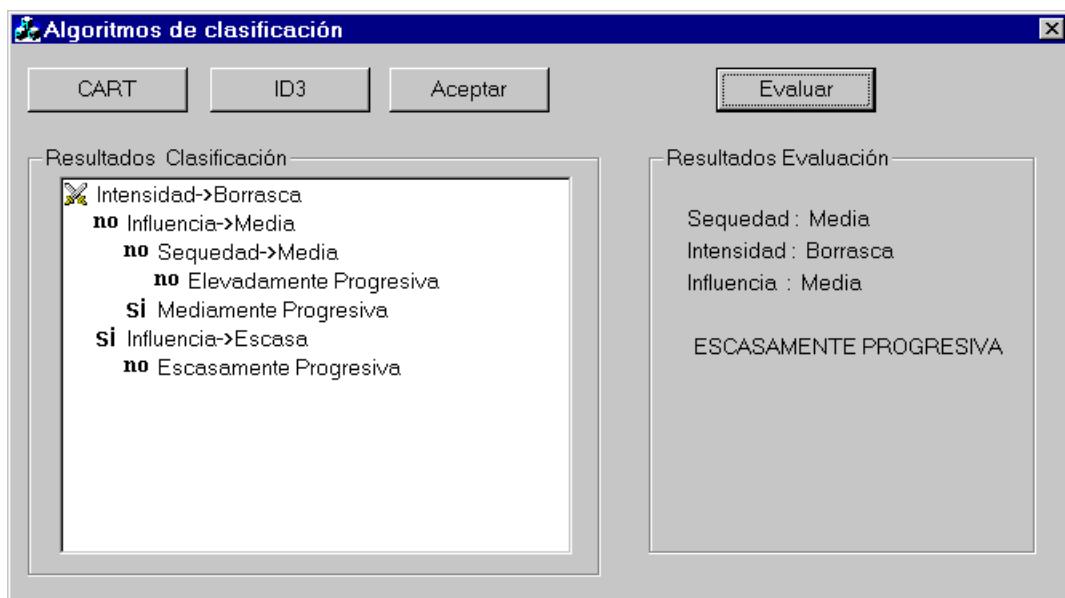


Figura 39: Interfaz y ejemplo de CART con poda.

1.6. Resumen del proceso de Adquisición de Conocimiento y resultados.

1.6.1. Adquisición de conocimiento inicial.

Se han usado técnicas habituales de adquisición de conocimiento: Entrevistas, cuestionarios, análisis estructural de textos ... Se ha logrado:

- Una definición del dominio exhaustiva y precisa.

- Una caracterización clara de las metas y objetivos del sistema.
- Información heurística para operar en diferentes fases de la adquisición automática de conocimiento.
- Un patrón de crecimiento en la evolución de la siniestralidad de los incendios forestales, donde el *primer sector* representa un lento crecimiento de la siniestralidad en los días inmediatamente posteriores al periodo de lluvias. El *segundo sector* expresa un alto crecimiento de la siniestralidad, especialmente en cuanto al número de incendios y en el *tercero* se quiere representar una estabilización en cuanto al número, pero un progresivo aumento en la peligrosidad de los fuegos.

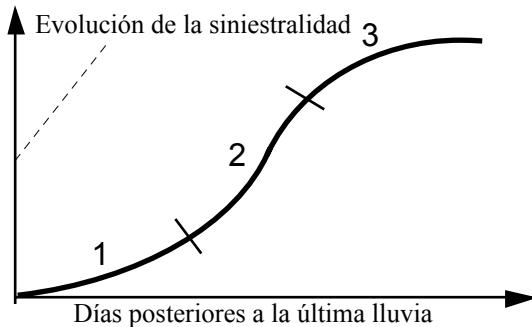


Figura 40.

1.6.2. DCPB, Descubrimiento de Conocimiento Prototípico Borroso.

Partiendo de la idea general del proceso *KDD* (*Knowledge Discovery in Databases*), se extiende y precisa, buscando no conocimiento en general, sino *Prototipos de ciclos* de comportamiento, entendiendo por prototipos la caracterización de *Categorías Prototípicas Borrosas* presentada. Las fases de DCPB son las siguientes:

1. *Selección*: Aplicando el conocimiento del dominio y el conocimiento relevante *a priori*, teniendo en cuenta los objetivos del proceso global de DCPB se ha creado una tarjeta de datos que toma como conjunto de partida una base de datos relacional que contiene aproximadamente 12.000 incendios ocurridos en Galicia durante los años 1991 y 1992. Se han seleccionado los 3.204

correspondientes a la provincia de Lugo, de los años 1991 y 1992. También se han separado por comarcas y se han eliminado campos no relevantes.

2. *Preproceso*: Se han limpiado los datos, se han eliminado los ruidos, se han manejado los campos vacíos, datos perdidos, valores desconocidos o por defecto y la evolución de datos. Se han aplicado técnicas estándar de bases de datos.
3. *Transformación*: Se ha reducido el número de variables. Se han localizado ciclos de incendios. Se han buscado formas útiles para expresar los datos dependiendo del uso posterior que se les va a dar y de los objetivos del sistema. Se han usado el conocimiento experto y técnicas de transformación e informes en bases de datos. También se ha utilizado una técnica de *clustering* jerárquico con lo que se ha conseguido la ordenación y clasificación de los ciclos según la evolución de la siniestralidad.
4. *Data Mining*: Se han elegido los algoritmos de *Data Mining*. Se han tomado decisiones acerca del modelo que se deriva del algoritmo de *Data Mining* elegido (resumen de datos). Se han buscado patrones de interés. Se han generado los prototipos de evolución de la siniestralidad en base a los ciclos representativos de cada una de las clases, y el resultado ha sido el siguiente:

	Sector 1	Sector 2	Sector 3
Media de días:	1	4	2
Media de incendios/día:	1	2	4
Mínimo de inc/día:	1	1	1
Máximo de inc/día:	1	4	9
Nº especialistas/día:	1	2	6
Nº de brigadas/día:	2	2	5
Nº de medios/día:	1	1	3

Tabla 63: Prototipo “Siniestralidad Escasamente Progresiva”.

	Sector 1	Sector 2	Sector 3
Media de días:	6	2	6
Media de incendios/día:	3	4	6
Mínimo de inc/día:	1	1	1
Máximo de inc/día:	8	11	19
Nº especialistas/día:	4	6	11
Nº de brigadas/día:	5	6	12
Nº de medios/día:	4	5	11

Tabla 64: Prototipo “Siniestralidad Medianamente Progresiva”.

	Sector 1	Sector 2	Sector 3
Media de días:	8	13	15
Media de incendios/día:	9	12	19
Mínimo de inc/día:	1	1	1
Máximo de inc/día:	31	53	85
Nº especialistas/día:	8	16	17
Nº de brigadas/día:	13	21	33
Nº de medios/día:	5	13	18

Tabla 65: Prototipo “Siniestralidad Altamente Progresiva”.

1.6.3. Representación de los Prototipos Borrosos.

Partiendo de la noción de Categorías Prototípicas Borrosas, estos prototipos de evolución de la siniestralidad han sido representados como números borrosos triangulares, con el fin de poder evaluar casos reales.

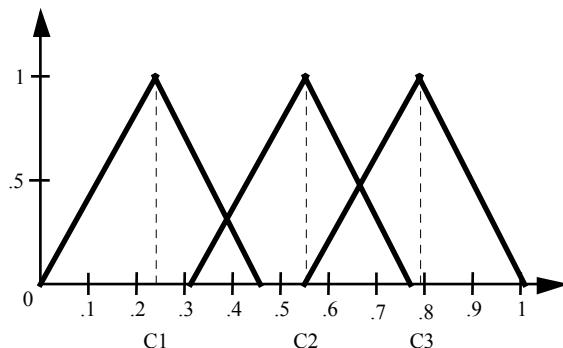


Figura 41: Representación de los prototipos.

1.6.4. Temas abiertos.

Para completar y hacer más eficiente el proceso DCPB aquí descrito, la investigación debe orientarse hacia los siguientes aspectos:

- *Conjuntos de datos de grandes dimensiones* (por ejemplo información geográfica) cuyo tratamiento puede orientarse hacia algoritmos muy eficientes, métodos de aproximación, procesamiento más eficiente (por ejemplo paralelismo masivo), etcétera.
- *Comprensibilidad de los prototipos* de datos, para que los patrones descubiertos sean más comprensibles por los usuarios, puede incluir representaciones gráficas, estructuras de reglas, lenguaje natural y otras muchas cosas.
- Tratamiento de los *datos perdidos*, por ejemplo con valores por defecto.
- Gestión de datos y conocimiento que *cambia muy rápidamente*.
- *Integración* con otros sistemas de las conclusiones obtenidas.
- *Datos no-estándar*: Multimedia, datos orientados a objetos, por ejemplo la información suministrada por satélites meteorológicos o geográficos, etcétera.
- Uso de *algoritmos de clasificación y clustering de última generación*, por ejemplo, clasificadores múltiples, agentes inteligentes ...

2. Matching Caso Real – Categorías Prototípicas Deformables Borrosas. Aplicación a la Predicción de incendios forestales.

Se presenta la aplicación a la predicción en los incendios forestales de un nuevo modelo que se apoya en estos tres pilares:

- El concepto de Prototipo Deformable de Bremermann.
- El razonamiento no-monótono con marcos.
- El razonamiento borroso no-monótono, utilizando las funciones propuestas por Castro y Trillas.

El modelo propuesto consiste en lo siguiente:

1. Aproximar (deformar) las Categorías Prototípicas Borrosas descubiertas mediante DCPB para que describan la situación cíclica (evolutiva) real.

2. Representar mediante marcos la definición de la situación cíclica real resultante.
3. Equiparar la situación diaria con estos marcos que describen la situación cíclica. Reajustar el modelo evolutivo si la confirmación diaria se desvía. Se utilizan los descritos en el Apéndice 1:
 - Razonamiento sobre *taxonomías prototípicas*.
 - Razonamiento temporal.
 - Razonamiento borroso no-monótono.

2.1. *Aproximación (deformación) de las Categorías Prototípicas Deformables Borrosas al ciclo real.*

Es necesario tener en cuenta que entre diferentes ciclos existen diferencias que vienen dadas por el cambio del valor de los factores de ciclo. Por tanto, es necesario que el día del inicio del ciclo (primer día sin lluvias) se genere un nuevo ciclo a partir de su afinidad con las Categorías Prototípicas Deformables Borrosas. Para ello se pregunta al usuario el valor real de los factores de ciclo. Su combinación dará unos grados de compatibilidad con las Categorías Prototípicas Deformables Borrosas, en particular los que interesan son aquellos que dan un valor de compatibilidad no nulo.

Para encontrar la/las Categorías Prototípicas Deformables Borrosas afines se establece el siguiente proceso:

1. Combinar los valores seleccionados por el usuario para el ciclo en curso (r).

M: operador de combinación Media.

f_i : valor de un factor. $i=1$ para el factor "Sequedad estacional", etc.

p_i : prototipo. $i=1$ para el prototipo "Siniestralidad escasamente progresiva", etc.

$$M(f_1, f_2, f_3) = r$$

2. Hallar los grados de compatibilidad con cada prototipo: $\mu_{p1}(r)$, $\mu_{p2}(r)$, $\mu_{p3}(r)$.
3. La/las Categorías Prototípicas Deformables Borrosas afines son las que tienen un grado de compatibilidad no nulo con r : los p_i tal que $\mu_{pi}(r) = (\mu_{p1}(r), \mu_{p2}(r), \mu_{p3}(r)) \neq 0$.

Ahora se podría adoptar una solución menos fina, consistente en asimilar el comportamiento del ciclo actual al de su prototipo más afín. Pero de lo que se trata es de modificar la/las Categorías Prototípicas Deformables Borrosas más afines hacia el ciclo que ahora se inicia. Ejemplo:

El usuario el primer día del ciclo introduce los siguientes valores: {Sequedad: Media/0.45, LLuvia: Tormenta/0.5, Influencia social: Mucha/0.85}. Lo cual resulta $M = 0.6$.

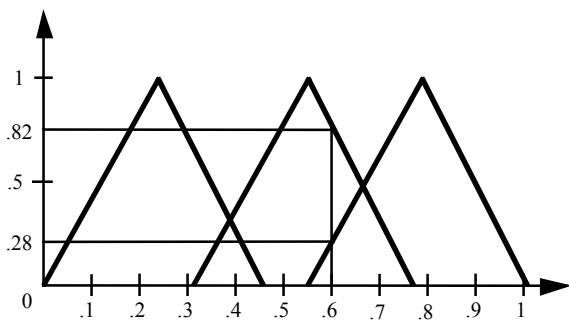


Figura 42.

Como se puede observar en la figura 42, los grados de compatibilidad son los siguientes:

$$\mu_{p1}(r) = 0$$

$$\mu_{p2}(r) = 0.82$$

$$\mu_{p3}(r) = 0.28$$

De lo que se deduce que las Categorías Prototípicas Deformables Borrosas más afines son p_2 y p_3 (“Siniestralidad medianamente progresiva” y “Siniestralidad altamente progresiva”).

Como anteriormente se ha dicho es necesario asimilar el comportamiento del ciclo actual al de su/sus Categorías Prototípicas Deformables Borrosas afines. Para lo

cual es preciso modificar la definición que se ha obtenido mediante DCPB de las Categorías Prototípicas Deformables Borrosas, evidentemente esta modificación debe realizarse en base al grado de compatibilidad con las Categorías Prototípicas Deformables Borrosas.

Aplicando la combinación lineal descrita en el concepto de *Categorías Prototípicas Deformables Borrosas*, la definición del ciclo real actual será la siguiente:

$$C_{real}(w_1 \dots w_n) = |\sum \mu p_i(v_1 \dots v_n)|$$

Donde:

C_{real} Ciclo real actual.

$(w_1 \dots w_n)$ Parámetros que describen el Ciclo real actual.

μp_i Grados de compatibilidad con las Categorías Prototípicas Deformables Borrosas distintos de 0.

$(v_1 \dots v_n)$ Parámetros de éstas Categorías Prototípicas Deformables Borrosas.

Para el ejemplo, con $\mu p_2 = 0.82$ y $\mu p_3 = 0.28$, la deformación de las Categorías Prototípicas Deformables Borrosas afines hacia el ciclo real es la siguiente:

$$C_{real}(v_1 \dots v_n) = 0.82 * \begin{pmatrix} 6 & 2 & 6 \\ 3 & 4 & 6 \\ 1 & 1 & 1 \\ 8 & 11 & 19 \\ 4 & 6 & 11 \\ 5 & 6 & 12 \\ 4 & 5 & 11 \end{pmatrix} + 0.28 * \begin{pmatrix} 8 & 13 & 15 \\ 9 & 12 & 19 \\ 1 & 1 & 1 \\ 31 & 53 & 85 \\ 8 & 16 & 17 \\ 13 & 21 & 33 \\ 5 & 13 & 18 \end{pmatrix}$$

Con lo que la definición del Ciclo real actual queda de la siguiente forma:

	Sector 1	Sector 2	Sector 3
Media de días:	7	5	9
Media de incendios/día:	5	7	10
Mínimo de inc/día:	1	1	1
Máximo de inc/día:	15	24	39
Nº especialistas/día:	6	9	14
Nº de brigadas/día:	8	11	19
Nº de medios/día:	5	8	14

Tabla 66: Definición del ciclo real para $\mu p_2 = 0.82$ y $\mu p_3 = 0.28$.

Recapitulando brevemente, se comienza determinando las Categorías Prototípicas Deformables Borrosas “ideales” de progresión de la siniestralidad; se califican como ideales ya que son paradigmas de progresión para cualquier ciclo. Lo que la tabla anterior muestra es la definición del ciclo actual a partir de la modificación de las Categorías Prototípicas Deformables Borrosas afines. Dicho de otra forma, se interpreta la situación actual a partir de la deformación de paradigmas previamente establecidos.

2.2. *Aproximación del ciclo real a la situación diaria (Predicción).*

La previsión establecida sobre el ciclo actual sería valida si los factores diarios no tuviesen peso alguno, cosa que no ocurre; más bien sucede que el ciclo no debería modificarse salvo que hubiese un cambio inesperado en la situación diaria. Por ejemplo, un aumento brusco de la sequedad local debido a vientos secos y constantes. Por ello, se establece una posible corrección del ciclo sobre la base de los factores diarios.

2.2.1. *Factores diarios.*

Los factores diarios que manejan los expertos y que se usarán en el sistema se reducen a los siguientes:

- A. **Número de días después de la última lluvia** (valor no borroso).
- B. **La serie de incendios de los días anteriores** (valor no borroso).

- C. **Virulencia de los siniestros del día anterior.** El usuario dispone de {Extrema/0.95, Alta/0.8, Media/0.5, Baja/0.2, Nula/0}.
- D. **Sequedad local.** Vendría dada por la probabilidad de ignición de los combustibles en la comarca que se estudia. El usuario la recibe diariamente en porcentaje, aunque el sistema asocia cada posible entrada (probabilidad de ignición) con una etiqueta y un grado, de la siguiente forma: {Extrema(60%)/1, Muy Alta(50%)/0.85, Alta(30%)/0.6, Media(40%)/0.4, Baja(20%)/0.25, Muy baja(10%)/0.05}.
- E. **Influencia social diaria.** Este factor debe recoger hechos tales como el que exista una mayor tendencia a provocar incendios durante el fin de semana, o también una mayor tendencia cuando hay una fiesta local. El usuario dispone de {Alta/0.9, Baja/0.6}. Si no hay ninguna influencia social conocida se ignora este valor.
- F. **Previsión de lluvias.** La experiencia demuestra que cuando las previsiones meteorológicas en los medios de comunicación predicen lluvias a corto plazo (en uno o dos días), el número de incendios provocados disminuye sustancialmente. Se puede comprobar en el anterior análisis detallado de ciclos. Esto explica que en muchos ciclos haya una caída acusada en los dos últimos días del ciclo. Por ello este factor será considerado como un valor booleano sí/no.

2.2.2. Definición del ciclo real.

No es necesario para modificar el ciclo real realizar una combinación de factores, tal y como se hizo en la modificación del Prototipo Conceptual. La razón de esta diferencia se encuentra en el conocimiento experto adquirido, el cual señala que en la modificación diaria no es pertinente una aproximación gradual. En la modificación diaria el motor del cambio son determinados valores de algunos de los factores, no la combinación de todos los factores. La modificación de las Categorías Prototípicas Deformables Borrosas era esencialmente cuantitativa, sin embargo la modificación del ciclo real es más bien cualitativa.

Todos los factores diarios, salvo la influencia social y la previsión de lluvias, ayudan a validar el buen comportamiento del ciclo real preestablecido. Una disparidad muy acusada con los factores previstos en el ciclo real llevaría a modificaciones, por ejemplo, un cambio prematuro de sector. La influencia social simplemente ayuda a avisar al usuario de que debe tomar las medidas necesarias en esa zona, por ejemplo, asignar más medios. También sirve de explicación de un aumento inusual del número de incendios en un momento determinado del ciclo (máximo local).

Un nuevo ciclo comienza cuando el usuario introduce los factores de ciclo, es decir, cuando deja de llover. El marco (*frame*) de referencia para establecer la predicción es el que se muestra en la tabla n:

Marco: Evolución de la siniestralidad para el ciclo real.			
Slot	Descripción/Valores	Inform. al Usuario	Nº
Día (Fecha)	Fecha real de cada uno de los días del ciclo	Se Muestra	1
Nº de incendios I	Distribución incremental por sector, de forma continua en los tres sectores para que el usuario observe una progresión	Se Muestra	2
Nº de incendios II	Distribución incremental por sector, local para cada sector, para poder descubrir en que día del sector se está, si no ha habido un cambio de sector	NO se Muestra	3
Media/Máximo de Incendios	Información de la [Media/Máximo] del número de incendios esperado para el día, para cada sector	Se Muestra	4
Parámetro de comparación del número de incendios	Se utilizará para comparar la situación diaria con el ciclo. Su valor es (Media+Máximo)/2 para cada uno de los ciclos	NO se Muestra	5
Número de especialistas	Distribución incremental por sector, normalizando la distribución en los intervalos: Sector 1: [Valor del número de especialistas/día del sector 1 del ciclo real ± distancia al valor medio entre el del sector 1 y el 2] Sector 2: [Valor del número de especialistas/día del sector 2 del ciclo real ± distancia al valor medio entre el del sector 1 y el 2]	Se Muestra	6

	Sector 3: [Valor del número de especialistas/día del sector 3 del ciclo real ± distancia al valor medio entre el del sector 2 y el 3]		
Intervalo y distribución de especialistas	Distribución no continua construida a partir de los intervalos cuya disparidad determinará un cambio de sector: Sector 1: [Valor del número de especialistas/día del sector 1 del ciclo real ± distancia al valor del segundo sector] Sector 2: [Valor del número de especialistas/día del sector 2 del ciclo real ± distancia al valor del primer sector] Sector 3: [Valor del número de especialistas/día del sector 3 del ciclo real ± distancia al valor del segundo sector]	NO se Muestra	7
Número de brigadas	Distribución incremental por sector, normalizando la distribución en los intervalos: Sector 1: [Valor del número de brigadas/día del sector 1 del ciclo real ± distancia al valor medio entre el del sector 1 y el 2] Sector 2: [Valor del número de brigadas/día del sector 2 del ciclo real ± distancia al valor medio entre el del sector 1 y el 2] Sector 3: [Valor del número de brigadas/día del sector 3 del ciclo real ± distancia al valor medio entre el del sector 2 y el 3]	Se Muestra	8
Intervalo y distribución de brigadas	Distribución no continua construida a partir de los intervalos cuya disparidad determinará un cambio de sector: Sector 1: [Valor del número de brigadas/día del sector 1 del ciclo real ± distancia al valor del segundo sector] Sector 2: [Valor del número de brigadas/día del sector 2 del ciclo real ± distancia al valor del primer sector] Sector 3: [Valor del número de brigadas/día del sector 3 del ciclo real ± distancia al valor del segundo sector]	NO se Muestra	9
Número de medios mecánicos	Distribución incremental por sector, normalizando la distribución en los intervalos:	Se Muestra	10

	Sector 1: [Valor del número de medios/día del sector 1 del ciclo real ± distancia al valor medio entre el del sector 1 y el 2] Sector 2: [Valor del número de medios/día del sector 2 del ciclo real ± distancia al valor medio entre el del sector 1 y el 2] Sector 3: [Valor del número de medios/día del sector 3 del ciclo real ± distancia al valor medio entre el del sector 2 y el 3]		
Intervalo y distribución de medios mecánicos	Distribución no continua construida a partir de los intervalos cuya disparidad determinará un cambio de sector: Sector 1: [Valor del número de medios/día del sector 1 del ciclo real ± distancia al valor del segundo sector] Sector 2: [Valor del número de medios/día del sector 2 del ciclo real ± distancia al valor del primer sector] Sector 3: [Valor del número de medios/día del sector 3 del ciclo real ± distancia al valor del segundo sector]	NO se Muestra	11
Intervalos de virulencia de los incendios del día anterior	Los intervalos cuya disparidad determinará un cambio de sector son los siguientes: Sector 1: [Baja/0.2, Nula/0] Sector 2: [Alta/0.8, Media/0.5, Baja/0.2] Sector 3: [Extrema/0.95, Alta/0.8]	NO se Muestra	12
Intervalos de Sequedad local	Los intervalos cuya disparidad determinará un cambio de sector son los siguientes: Sector 1: [Media(40%)/0.4, Baja(20%)/0.25, Muy baja(10%)/0.05] Sector 2: [Alta(30%)/0.6, Media(40%)/0.4] Sector 3: [Extrema(60%)/1, Muy Alta(50%)/0.85, Alta(30%)/0.6]	NO se Muestra	13
Influencia social diaria	Valor por defecto: 0 (No hay influencia social relevante)	NO se Muestra	14
Previsión de lluvias	Valor por defecto: 0 (No hay previsión de lluvias)	NO se Muestra	15

Tabla 67: Marco Evolución de la siniestralidad para el ciclo real.

El marco que servirá para hacer la equiparación con el de “Evolución de la siniestralidad para el ciclo real” (sus *slots* están representados en este nuevo marco con números) y que determinará los posibles cambios de sector o incidencias diarias puntuales es el de “Situación diaria”, que queda reflejado en la tabla 68:

Marco: Situación diaria		
Slot	Descripción/Valores	Id.
Número de incendios del día anterior	El valor se toma de la Base de Datos. Se compara con 5 para determinar el sector. Se compara con 3 para determinar el día. Se mostrará al usuario para que tenga una referencia.	A
Número de especialistas del día anterior	El valor se toma de la Base de Datos. Se compara con los intervalos de 7 para determinar el sector. Se compara con las distribuciones de 7 para determinar el día.	B
Número de brigadas del día anterior	El valor se toma de la Base de Datos. Se compara con los intervalos de 9 para determinar el sector. Se compara con las distribuciones de 9 para determinar el día.	C
Número de medios del día anterior	El valor se toma de la Base de Datos. Se compara con los intervalos de 11 para determinar el sector. Se compara con las distribuciones de 11 para determinar el día.	D
Virulencia de los incendios del día anterior	El valor es introducido diariamente por el usuario Se compara con 12 para determinar el sector. Los posibles valores son: {Extrema/0.95, Alta/0.8, Media/0.5, Baja/0.2, Nula/0}.	E
Sequedad Local	El valor es introducido diariamente por el usuario Se compara con 13 para determinar el sector. Los posibles valores son: {Extrema(60%)/1, Muy Alta(50%)/0.85, Alta(30%)/0.6, Media(40%)/0.4, Baja(20%)/0.25, Muy baja(10%)/0.05}.	F
Influencia social diaria	El valor es introducido diariamente por el usuario Se compara con 14 para determinar aspectos específicos del comportamiento diario. Los posibles valores son: {Alta/0.9, Baja/0.6}.	G
Previsión de lluvias	El valor es introducido diariamente por el usuario Se compara con 15 para determinar aspectos específicos del comportamiento diario. Los posibles valores son: {Sí, No}.	H
Número de días desde la última lluvia	Se mostrará al usuario para que tenga una referencia.	-

Tabla 68: Marco Situación diaria.

*2.2.3. Equiparación de la situación diaria con el ciclo real (Pattern Matching).
Modificaciones del ciclo real (si son necesarias).*

La equiparación de la situación diaria con el ciclo real se realiza en dos fases, una para determinar en qué sector se está y si hay un cambio de sector, la segunda para determinar con exactitud en qué día del sector se está y poderle mostrar al usuario la predicción para los próximos días.

1ª Fase: Determinación del sector.

Para determinar en qué sector se encuentra la evolución de la siniestralidad, teniendo en cuenta que no es posible un retroceso en el sector (como se ha visto anteriormente), se procede llevando a cabo las siguientes ocho operaciones:

1. Comparar A (Número de incendios del día anterior) con 5 (Parámetro de comparación del número de incendios). El sector será aquel que refleje una distancia menor entre ambos valores.
2. Comparar B (Número de especialistas del día anterior) con 7 (Intervalo de especialistas). Si hay disparidad (B no pertenece al intervalo), cambiar al siguiente sector que cumpla la pertenencia.
3. Comparar C (Número de brigadas del día anterior) con 9 (Intervalo de brigadas). Si hay disparidad (C no pertenece al intervalo), cambiar al siguiente sector que cumpla la pertenencia.
4. Comparar D (Número de medios del día anterior) con 11 (Intervalo de medios). Si hay disparidad (D no pertenece al intervalo), cambiar al siguiente sector que cumpla la pertenencia.
5. Comparar E (Virulencia de los incendios del día anterior) con 12 (Intervalos de virulencia de los incendios del día anterior). Si hay disparidad (E no pertenece al intervalo), cambiar al siguiente sector que cumpla la pertenencia.
6. Comparar F (Sequedad Local) con 13 (Intervalos de Sequedad local). Si hay disparidad (F no pertenece al intervalo), cambiar al siguiente sector que cumpla la pertenencia.

7. Comparar G (Influencia social diaria) con 14 (Influencia social diaria). Si hay disparidad (G tiene un valor distinto de 0), llevar a cabo las siguientes acciones:

- Mostrar los siguientes mensajes al usuario:

Para BAJA: “Hoy podría haber más incendios de los previstos debido a causas sociales de baja intensidad. Por favor, preste especial atención a las zonas afectadas y destine más recursos a las mismas”.

Para ALTA: “Hoy podría haber más incendios de los previstos debido a causas sociales de alta intensidad. Por favor, preste especial atención a las zonas afectadas y destine más recursos a las mismas”.

- Estancar la evolución y considerar ese día como un máximo local.

8. Comparar H (Previsión de lluvias) con 15 (Previsión de lluvias). Si hay disparidad (H tiene un valor distinto de 0), llevar a cabo las siguientes acciones:

- Mostrar el siguiente mensajes al usuario:

“Debido al anuncio de lluvias por los servicios meteorológicos se podría producir un descenso en la intencionalidad de los incendios. Aunque ocurran menos alarmas, no descuida la vigilancia hasta que realmente llueva”.

- Aunque haya previsión de lluvias continuar normalmente con la evolución del ciclo (es posible que decrezca el número de incendios, lo que provocaría un retroceso en los días del ciclo).

2^a Fase: Determinación del día dentro del sector.

Para determinar el día dentro del sector en el que se debe posicionar la evolución, se calcula la distancia en valor absoluto (Valores normalizados) con cada uno de los días del ciclo. Se utilizará la función de similaridad que ha sido usada para el *clustering jerárquico*, sea:

$$d(u, w) = \sum_i |x_i - y_i|$$

Se comparará el vector [A (Nº de incendios del día anterior), B (Número de especialistas del día anterior), C (Número de brigadas del día anterior), D (Número de medios del día anterior)] con el vector [3 (Nº de incendios II), 7 (Distribución de especialistas), 9 (Distribución de brigadas), 11 (Distribución de medios)] de cada día del sector.

El sistema desarrollará toda la predicción posicionándose en el día más afín a la situación diaria real y generando a partir de ahí diaria y dinámicamente la evolución del ciclo real. Esta retroalimentación - comprobación diaria es la que provoca que en los primeros días de un ciclo el error en la predicción sea alto, y a medida que se avanza sea cada vez más precisa.

2.2.4. Implementación.

Al comenzar un nuevo ciclo real se deben introducir los factores de ciclo ya descritos; a partir de estos datos se manipulan de la siguiente manera:

Para la realización de estas tareas se contará con una lista de prototipos que contendrá los datos sobre los prototipos conceptuales procedentes de la adquisición de conocimiento. El producto o salida de este primer proceso es un Ciclo completo de días con datos de varias naturalezas: distribuciones continuas para el número de incendios y de recursos para cada día, distribuciones locales para cada sector definidas dentro de un intervalo y valores por defecto de los factores diarios no cuantitativos.

1. Inicialización de los prototipos reales con los datos procedentes de la adquisición de conocimiento.
2. Obtención de los factores de ciclo.
3. Combinación de los valores de los factores de ciclo mediante una media aritmética.
4. Calculo de los grados de compatibilidad del valor actual con los prototipos.

Para cada Número Borroso representante de un prototipo:

Calcular el Grado de Afinidad con valor combinado (fórmulas normales).

5. Construcción del prototipo real que se adapta a la situación.

5.1 Si se posee afinidad a dos de los prototipos

5.1.1. Combinar todos los valores de los dos prototipos con los que tiene afinidad mediante una media aritmética ponderada.

5.1.1.1 Para cada uno de los sectores Combinar N° de días del sector, N° medio de incendios, N° máximo de incendios, N° de recursos según su tipo (especialistas, brigadas o vehículos).

5.2 Si no:

5.2.1 Si sólo tiene afinidad con el prototipo “escasamente progresiva” o con prototipo “medianamente progresiva”, Entonces Adaptar multiplicando todos los valores por el grado de afinidad con el prototipo.

5.2.2 Si sólo tiene afinidad con el prototipo “altamente progresiva”, entonces multiplicar todos los valores del prototipo por el valor (2- grado de afinidad con el prototipo).

6. Modificación Diaria del Ciclo Real.

Construcción del ciclo real.

Las tareas numeradas 1-5 se realizan sobre la clase Lista de Prototipos que posee 3 Prototipos, mientras que la sexta tarea se realiza ya sobre el verdadero objeto/marco del ciclo real construyéndose de forma completa dando lugar a una primera aproximación de todos los datos del ciclo.

En sí, el marco objeto del ciclo real se divide en dos parte que tienen mucho en común pero tienen una finalidad distinta, una parte será la conformada por los datos que se mostrarán al usuario mientras que la otra parte se utilizará para realizar el análisis de los datos para confrontarlos con los datos reales que se vayan produciendo y poder así adaptar el ciclo construido primeramente a la situación real que se vaya produciendo.

Construcción de las listas de muestra: Nº de incendios previsibles, media y máximo de incendios (valores fijos dentro de cada sector), Nº de especialistas, patrullas y medios mecánicos previsiblemente ocupados.

1. Inicialización de las distribuciones:

1.1 INCENDIOS: determinación de los incendios del primer día del ciclo:

- DIF = Diferencia entre el máximo y la media de incendios prevista para el primer sector
- FUEGOSDIA = Media de incendios del primer sector - DIF OR 1 (si el valor anterior es menor o igual que cero).

1.2 RECURSOS: Determinación de los recursos (según su tipo) del primer día del ciclo:

- DIF = Diferencia entre el promedio de recursos del tipo correspondiente entre el sector uno y el sector dos.
- Recursos (según tipo) del primer día = Promedio de recursos esperado - DIF OR 1 si el valor anterior es menor que cero.

2. Para cada sector:

2.1 Readaptar parámetros locales según el sector:

2.1.1 INCENDIOS

$$\begin{aligned} \text{DIFFUEGOS} &= \text{Sect[ii].m_nMaxFuegosDia} - \\ &\quad \text{FUEGOSDIA}; \end{aligned}$$

FUEGOSDIARIOS = DIFUEGOS / Días del Sector OR 1 si el resultado es igual a 0

2.1.2. RECURSOS

RECURSOSDIARIOS son los recursos que hay que aumentar cada día.

- a) Si el sector es el último

DIFRECURSO = Diferencia entre el promedio del recurso del sector anterior y el actual.

b) Si no

DIFRECURSO = Diferencia entre el promedio del recurso del sector.

c) RECURSOSDIARIOS = ((Promedio del sector + DIFRECURSO) - recursosdia) / Días del sector.

2.2 Para cada uno de los días dentro del sector

2.2.1 Insertar los datos de los días en la lista. Para ello construir CDias(Sect[ii],ii,fuegosdia,specdia,patrdia,mechdia);

2.2.2 Actualización de datos distribuidos

- Incrementar número incendios (estancar si sobrepasa al máximo)

$$\text{FUEGOSDIA} = \text{FUEGOSDIA} + \text{FUEGOSDIARIOS}$$

- Incrementar número de recursos (Especialistas, Brigadas y Vehículos)

$$\text{RECURSOS DIARIOS} = \text{RECURSOS DIA} + \text{RECURSOSDIARIOS}$$

2.3 Almacenar los días del sector

3. Inicializar datos generales

3.1 Número de días del ciclo = suma de los días de los tres sectores.

3.2 Sector_actual = 0 y Dia_Actual del Ciclo = 0

Complejidad del algoritmo de construcción del ciclo real: Factores determinantes: F (número de factores de ciclo), P (nº de prototipos), S (nº de sectores de cada prototipo), N (nº de datos de cada sector), D (nº de días del ciclo).

Combinación de los factores de ciclos (F).

Cálculo de la afinidad a los prototipos (P).

Cálculo del prototipo real: ($S * P * N$ que son números pequeños).

Cálculo del ciclo real de días: $k*D$.

Cálculo de los datos sectoriales: $S*N$.

Cálculo del ciclo de análisis: $k*D$.

Complejidad total: Max (F, P, $S*P*N$, $k * D$, $S*N$, $k*D$).

Posteriormente se generan todos los datos correspondientes al análisis, que constan de datos de dos tipos: Unas distribuciones locales para cada sector que se corresponden con cada una de las distribuciones continuas que se muestran y que ya se han construido y unos datos propios para cada sector entre los que se incluyen los intervalos de validación de todos los factores tanto cuantitativos (N^o de incendios, N^o de especialistas, patrullas y vehículos) como cualitativos (sequedad local y virulencia de los incendios) así como valores escalares como el indicador del promedio de incendios y los valores por defecto de los factores diarios que determinan casos especiales (influencia social y previsión de lluvias).

1. Inicialización de datos sectoriales

1.1 Inicialización de datos cuantitativos

Para cada sector ss hacer

$$\begin{aligned} 1.1.1 \quad & \text{Número Indicador de Incendios} = \\ & \text{Sect}[ss].\text{MediaIncendiosDiarios} + \\ & (\text{Sect}[ss].\text{MaximoIncendiosDiarios}) / 2 \end{aligned}$$

1.1.2 Indicadores de recursos (según su tipo)

a) Si el sector es el último

$$\begin{aligned} \text{DIFRECURSO} &= \text{Sect}[ss].\text{PromedioRecurso} - \text{Sect}[ss-1]. \\ &\text{PromedioRecurso} \end{aligned}$$

$$\begin{aligned} b) \quad & \text{Si no DIFRECURSO} = \text{Sect}[ss+1].\text{PromedioRecurso} - \text{Sect}[ss]. \\ &\text{PromedioRecurso} \end{aligned}$$

c) Construir Intervalo Recursos: CIntervalo(Mínimo, máximo)

CIntervalo(Sect[ss].PromedioRecurso-DIFRECURSO,
Sect[ss].PromedioRecurso + DIFRECURSO)

1.2 Inicialización por defecto de los valores de pico

Sector[ss].m_InfSoc = false ;

Sector[ss].m_PrevLluv = false;

1.3 Inicialización de datos cualitativos

1.3.1 Virulencia de los incendios

Sector[0].m_IVirulencia = new CIntervalo(0.00,0.2);

Sector[1].m_IVirulencia = new CIntervalo(0.2,0.8);

Sector[2].m_IVirulencia = new CIntervalo(0.8,0.95);

1.3.2 Sequedad Local (Probabilidad de Ignición)

Sector[0].m_ISequedadLocal = new CIntervalo(0.05,0.4);

Sector[1].m_ISequedadLocal = new CIntervalo(0.4,0.6);

Sector[2].m_ISequedadLocal = new CIntervalo(0.6,1.00);

2 Construcción de las listas (distribuciones) de análisis

2.1 Inicializar los datos locales a cada sector (incendios y recursos del primer día del sector)

2.3.1 INCENDIOS

FUEGOSDIA son los incendios del día actual del sector (al inicializar el primero).

FUEGOSDIARIOS son los incendios que hay que aumentar cada día

DIFUEGOS = Máximo del sector - Media del sector

FUEGOSDIA = Media del sector - DIFFUEGOS OR 1 si el resultado anterior es negativo

FUEGOS DIARIOS = (Máximo del sector - FUEGOSDIA) / Días del sector

2.1.2 RECURSOS (según el tipo las variables son unas u otras)

RECURSOSDIA son los recursos de día actual del sector (al inicializar el primero).

RECURSOSDIARIOS son los recursos que hay que aumentar cada día

Si el sector es el último

DIFRECURSO = Diferencia entre el promedio del recurso del sector anterior y el actual

Si no

DIFRECURSO = Diferencia entre el promedio del recurso del sector

RECURSOSDIA = Promedio del sector actual - DIFRECURSO OR 1 si el resultado anterior es menor que cero.

RECURSOSDIARIOS = ((Promedio del sector + DIFRECURSO) - recursosdia) / Días del sector

2.2 Para cada uno de los días dentro del sector

2.2.1 Insertar los datos de los días en la lista. Para ello construir
CDias(Sector,ii,fuegosdia,specdia,patrdia,mechdia)

2.2.2 Actualización de datos distribuidos

- Incrementar número incendios (estancar si sobrepasa al máximo)

FUEGOSDIA = FUEGOSDIA + FUEGOSDIARIOS

- Incrementar número de recursos según su tipo (Especialistas, Patrullas y Vehículos) (estancar si sobrepasa al máximo)

RECURSOSDIA = RECURSOSDIA + RECURSOSDIARIOS

Como se puede observar la diferencia entre las distribuciones de análisis y las de muestra reside en que estas son continuas en cuanto al cambio de sector (el nuevo sector comienza en el último valor del sector anterior) mientras que el comienzo de cada sector en las distribuciones de análisis es calculado individualmente lo que da lugar a solapamientos entre los datos de los sectores.

Distribuciones mostrada y de análisis del ejemplo.

Estos son los datos que se muestran al usuario, donde primero se ve el día (fecha real), el número de incendios esperado, la media y el máximo de incendios esperados y los recursos necesarios esperados (especialistas, brigadas y medios mecánicos). Se puede observar que los días 8 y 13 se produce un cambio de sector.

Mar 01->[1](5,15):R(3, 5, 2)
Mar 02->[3](5,15):R(4, 6, 3)
Mar 03->[5](5,15):R(5, 7, 4)
Mar 04->[7](5,15):R(6, 8, 5)
Mar 05->[9](5,15):R(7, 9, 6)
Mar 06->[11](5,15):R(8,10, 7)
Mar 07->[13](5,15):R(9,11, 8)
Mar 08->[16](7,24):R(10,12, 9)
Mar 09->[18](7,24):R(12,14,11)
Mar 10->[20](7,24):R(14,16,13)
Mar 11->[22](7,24):R(16,18,15)
Mar 12->[24](7,24):R(18,20,17)
Mar 13->[25](10,40):R(20,22,19)
Mar 14->[27](10,40):R(21,23,20)
Mar 15->[29](10,40):R(22,24,21)
Mar 16->[31](10,40):R(23,25,22)
Mar 17->[33](10,40):R(24,26,23)
Mar 18->[35](10,40):R(25,27,24)

Mar 19->[37](10,40):R(26,28,25)

Mar 20->[39](10,40):R(27,29,26)

Mar 21->[40](10,40):R(28,30,27)

Los datos de análisis, ocultos para el usuario, con los que se comparará cada día son el día (numerado, no la fecha real), número de incendios (no con una distribución continua, sino con la de trabajo), intervalo de confianza de media y máximo de incendios (obsérvese que el máximo es mayor que el que se muestra para propiciar solapamientos más estables) y el número de recursos que previsiblemente se usarán (obsérvese que también son distintos):

Día 0->[1](5,15):R(3, 5, 2)

Día 1->[3](5,15):R(4, 6, 3)

Día 2->[5](5,15):R(5, 7, 4)

Día 3->[7](5,15):R(6, 8, 5)

Día 4->[9](5,15):R(7, 9, 6)

Día 5->[11](5,15):R(8, 9, 7)

Día 6->[13](5,15):R(9, 9, 8)

Día 7->[1](7,24):R(4, 1, 1)

Día 8->[6](7,24):R(6, 5, 4)

Día 9->[11](7,24):R(8, 9, 7)

Día 10->[16](7,24):R(10,13,10)

Día 11->[21](7,24):R(12,17,13)

Día 12->[1](10,40):R(9, 1, 1)

Día 13->[5](10,40):R(10, 5, 3)

Día 14->[9](10,40):R(11, 9, 5)

Día 15->[13](10,40):R(12,13, 7)

Día 16->[17](10,40):R(13,17, 9)

Día 17->[21](10,40):R(14,21,11)

Día 18->[25](10,40):R(15,22,13)

Día 19->[29](10,40):R(16,22,15)

Día 20->[33](10,40):R(17,22,17)

Adaptación al ciclo real en su transcurso.

Durante la duración del ciclo actual se deben tener en cuenta cada día los factores diarios para así adaptar la previsión ideal a las condiciones reales que se están dando durante los días del ciclo.

1. Obtención de los factores diarios procedentes de la Base de Datos: Número de incendios del día anterior, número de especialistas, patrullas y vehículos ocupados en estos incendios.
2. Obtención de los factores diarios cualitativos introducidos por el usuario como son la virulencia de los incendios del día anterior, la sequedad local presente (probabilidad de ignición) y los factores para casos especiales, la influencia social que se presenta en el día y la previsión de lluvias. Para ayudar al usuario se le muestran por pantalla algunos de los datos tomados de la BD como el número de incendios del día anterior y el número de días que ha pasado desde la última lluvia.

Con estos datos ADAPTAR el ciclo real (lo que se muestra) a las condiciones que se presentan en la realidad.

1. Si hay presencia de Influencia Social (baja o alta) se estanca la predicción del ciclo (avanzan solo las fechas) y se presenta un mensaje al efecto.
2. Si hay previsión de lluvias se estanca la predicción del ciclo (avanzan solo las fechas) y se presenta un mensaje al efecto.
3. Determinación del sector a partir del cual va a seguir el ciclo:

3.1 Se determina el sector de mayor afinidad comparando el índice de incendios de cada sector con el de incendios del día anterior. De esta forma se determinan los saltos de sector naturales por número de incendios:

Desde SectorActual hasta UltimoSector Hacer

Comparar IndiceIncendiosSector con NumeroIncendiosAyer

si la comparacion es la mejor actual almacenar sector como

MEJORSECTOR

3.2 Si (*MEJORSECTOR* \leftrightarrow *SECTORACTUAL*) entonces se produce un salto natural por incendios y no se tienen en cuenta los factores diarios: IR A 3.4

3.3 Si no hay salto por Número de INCENDIOS entonces

NOTA: Datos de la BD y el Usuario son: SequedadLocal, Virulencia, NumeroMecánicos, NumeroPatrullas, NumeroEspecialistas) los intervalos respectivos dentro del Sector[i] son: m_ISequedadLocal, m_IVirulencia, m_IMech, m_IPatrols, m_ISpec. Los métodos EstaEn del Intervalo devuelven true si el parámetro no supera al máximo del intervalo.

SALTO = NOT SectorActual.m_ISequedadLocal->EstaEn(SequedadLocal)
OR NOT SectorActual.m_IVirulencia->EstaEn(Virulencia) OR NOT
SectorActual.m_IMech->EstaEn(NumeroMecanicos) OR NOT
SectorActual.m_ISpec->EstaEn(NumeroEspecialistas) OR NOT
SectorActual.m_IPatrols->EstaEn(NumeroPatrullas).

o sea, SALTO será TRUE cuando hay alguna disonancia entre los valores del día y los que corresponden al sector Actual.

Si SALTO = TRUE y no se está en el último sector entonces el sector actual pasa a ser el siguiente.

3.4 Si se ha activado el SALTO

Redistribuir los datos de las listas del nuevo sector con los correspondientes de las listas de análisis

3.5 Hacer el *Matching* para determinar el día del sector más cercano al actual

3.5.1 Construir con los datos del día anterior un nuevo día

ULTDIA = CDIA(Datos del día anterior)

3.5.2 Para Cada DIA del SECTOR

3.5.2.1 DISTANCIA NORMALIZADA(*ULTDIA – DIA*)

3.5.2.2 Si el resultado es el mejor encontrado almacenarlo y seguir

3.5.3 Regresar el día más cercano (o el siguiente) MEJORDIA

3.6 Reajustar todas las listas de muestra poniendo MEJORDIA como primer día

Complejidad del algoritmo de readaptación del ciclo real: Factores determinantes: V (número de factores diarios), P (nº de prototipos), S (nº de sectores de cada prototipo), N (nº de datos de cada sector), D (nº de días totales del ciclo), D_i (días del sector i).

Matching sector: (S).

Comparación con valores diarios: (V).

Redistribución: Max (D_i) (i: 1...S).

*Matching día: Max(D_i)*N*c, (c: cte de cálculo).*

Complejidad total: Max(S, V, max(D_i), max(D_i)*N) = max(D_i)*N.

Este factor de complejidad habría que multiplicarlo por el nº de días que dure el ciclo. Suponiendo que la predicción indica correctamente el nº de días del ciclo entonces la complejidad total será: $D * \max(D_i) * N$ y como el máximo de D_i estaría acotado por D entonces $D^2 * N$.

Ejemplo

Con los datos iniciales de análisis del apartado anterior se le presentan al usuario los siguientes datos (situación original) antes de empezar un nuevo día del ciclo.

Mar 04->[3](5,15):R(4, 6, 3)

Mar 05->[5](5,15):R(5, 7, 4)

Mar 06->[7](5,15):R(6, 8, 5)

Mar 07->[9](5,15):R(7, 9, 6)

Mar 08->[11](5,15):R(8,10, 7)

Mar 09->[13](5,15):R(9,11, 8)

Mar 10->[16](7,24):R(10,12, 9)

Mar 11->[18](7,24):R(12,14,11)

Mar 12->[20](7,24):R(14,16,13)

Mar 13->[22](7,24):R(16,18,15)

Mar 14->[24](7,24):R(18,20,17)

Mar 15->[25](10,40):R(20,22,19)

Mar 16->[27](10,40):R(21,23,20)

Mar 17->[29](10,40):R(22,24,21)

Mar 18->[31](10,40):R(23,25,22)

Mar 19->[33](10,40):R(24,26,23)

Mar 20->[35](10,40):R(25,27,24)

Mar 21->[37](10,40):R(26,28,25)

Mar 22->[39](10,40):R(27,29,26)

Mar 23->[40](10,40):R(28,30,27)

Al comenzar el día el usuario introduce los datos de los factores diarios y se toman de la BD los datos de cantidades de incendios y recursos del día anterior.

Días desde última lluvia: 4

Incendios día anterior : 6

Especialistas: 3

Patrullas: 6

Medios Mecánicos: 6

Sequedad Local: 0.40

Virulencia: 0.50

Influencia Social Diaria: 0.00

Después de tratar estos datos y percibir un cambio de sector debido a factores diarios (Sequedad Local y Virulencia de los Incendios) la situación de predicción del día anterior se re adapta y presenta los siguientes resultados.

Mar 05->[11](7,24):R(8, 9, 7)

Mar 06->[16](7,24):R(10,13,10)

Mar 07->[21](7,24):R(12,17,13)

Mar 08->[25](10,40):R(20,22,19)

Mar 09->[27](10,40):R(21,23,20)

Mar 10->[29](10,40):R(22,24,21)

Mar 11->[31](10,40):R(23,25,22)

Mar 12->[33](10,40):R(24,26,23)

Mar 13->[35](10,40):R(25,27,24)

Mar 14->[37](10,40):R(26,28,25)

Mar 15->[39](10,40):R(27,29,26)

Mar 16->[40](10,40):R(28,30,27)

3. Redistribución de recursos.

El módulo de consejo (ayuda a la decisión), posee entre otras características, la de predicción de número e intensidad de incendios en una zona determinada y en base a esta predicción el consejo sobre los recursos a utilizar, su horario y situación. Además ayuda en la resolución del problema de reorganizar los recursos disponibles en una zona para lograr que ésta esté cubierta óptimamente (en el sentido de vigilancia y llegada rápida a cualquier punto de la zona donde se pueda originar un siniestro), cuando parte de los recursos que inicialmente mantenían el equilibrio se ven obligados a desplazarse y perder su disponibilidad debido a la necesidad de sofocar uno o varios incendios, o en casos muy especiales a otro tipo de causas (por ejemplo una avería mecánica).

Se presenta un método encaminado a la gestión eficaz y coherente con las soluciones que adoptarían los expertos, de la reorganización de medios en el sentido descrito.

3.1. *Presentación del problema.*

Como introducción al problema se describen las fases por las que pasa la distribución de medios:

1. **Colocación inicial de los recursos.** Como no todos los medios están activos siempre, cuando alguno comienza su actividad, el usuario decide su colocación en base a sus propios criterios. Esta colocación debe suponer una cobertura óptima de la zona. Habitualmente los recursos comienzan su actividad en puntos determinados que se denominarán bases.

2. **Inicio de incendios y asignación de medios.** Cuando se inicia algún incendio, el usuario envía los recursos que considera suficientes para controlarlo. La elección de los recursos enviados está determinada básicamente por su proximidad y el peligro potencial que represente el incendio.
3. **Distribución de los medios no asignados.** Como algunos medios han sido ocupados en la extinción del incendio o incendios iniciados, es conveniente alterar la situación inicial de los recursos no ocupados para lograr una cobertura lo más completa posible de la zona. Esta redistribución debe satisfacer una serie de características:
 - Debe ser lo más equilibrada posible, con un nivel máximo de cobertura, en cuanto a observación y rapidez en el acceso a todas las zonas.
 - Debe tener en cuenta las características específicas de cada recurso, por ejemplo es conveniente que las motobombas estén más próximas a sus bases que las brigadas.
 - El equilibrio no debe ser geométrico, ya que hay que tener en cuenta la existencia de zonas de importancia estratégica (especial valor ecológico, mayor peligrosidad potencial ...).
4. **Fin de un incendio y redistribución de los recursos.** Una vez se da por finalizado un incendio, los recursos que actuaban en él vuelven a estar libres y se presenta la necesidad de reorganizar la distribución de medios en la zona.

Esta parte del trabajo está encaminada a resolver el problema de distribución de medios no asignados (3).

3.2. *Dominio del problema.*

Se describen ahora las entidades cuya definición es necesaria para la aplicación del método propuesto para abordar la cuestión:

- **Zona o espacio de distribución.** Superficie de terreno donde se realiza la distribución de recursos. Se ha escogido como zona-tipo un ayuntamiento

(Lugo), como parte de una comarca administrativa compuesta por 6 ayuntamientos.

- **Puntos significativos.** Puntos geográficos de referencia relevantes por su importancia ecológica, calidad para la observación, peligrosidad potencial o nudo de comunicaciones. Los atributos de la entidad Punto significativo son los siguientes:

Objeto: Punto significativo		
Atributo	Valor	Observaciones
Posición	(x,y)	
Relevancia (R)	$R \in [0,1]$	Etiqueta lingüística dada por el experto.
Grado en que es cubierto por los recursos libres(a)	$a \in [0,1]$	Resultado de combinar la cobertura que ejercen sobre el punto todos y cada uno de los recursos libres.

Tabla 69: Definición de Punto Significativo.

- **Recursos libres.** Aquellos recursos no ocupados en la extinción de un incendio en un momento determinado. Su tabla O/A/V es la siguiente:

Objeto: Recurso libre		
Atributo	Valor	Observaciones
Posición	(x,y)	
Tipo	Motobomba/Brigada/ Helic.	
Movilidad (M)	$M \in [0,1]$	Etiqueta lingüística dada por el experto, que describe la movilidad del recurso.
Proximidad (P)	lista (P) $P = \{p_1, p_2, \dots, p_n\}$ ($p_i \in [0,1]$)	p_i es el grado de proximidad del recurso al Punto significativo s_i . Se obtiene a partir de la distancia en Km. mediante una Función de distribución.
Grados de cobertura a los respectivos Puntos significativos	lista (C)	Cada elemento de la lista identifica el grado en que un Punto es cubierto por este recurso, resultado de combinar la movilidad, la proximidad a cada punto y la relevancia de cada punto.
Nivel de cobertura general (G)	$G \in [0,1]$	Combinación de todos los elementos de C . Medida de lo bien que cumple su función cada recurso.

Tabla 70: Definición de Recursos Libres.

Se consideran de un modo general para conservar la sencillez del ejemplo, tres tipos de medios:

1. **Motobombas:** Camiones con una carga de agua, su movilidad es la más limitada entre estos recursos, deben alejarse lo menos posible de las bases.

2. **Brigadas de personal:** Grupos de personas con un vehículo todoterreno. Movilidad más amplia y flexible que la de las motobombas.

3. **Helicópteros de transporte:** Con personal equipado, utilizados como primer ataque rápido. Movilidad total.

3.3. *Método de distribución de los recursos no asignados.*

El método se compone de los siguientes pasos:

Paso 1. Determinación de las áreas que no están suficientemente cubiertas por los recursos, esta operación se subdivide en dos:

1. Para todos los Puntos significativos, determinar a (grado en que cada Punto significativo es cubierto por todos los recursos libres).
2. Se seleccionan todos los puntos cuyo a es menor que un cierto umbral dado y se incluyen en una lista I .

Paso2. Determinación de si es necesario mover algún recurso: Si I es una lista vacía significa que todas las áreas están suficientemente cubiertas y por tanto no es necesario mover ningún recurso y se termina el método. Si I no es vacía, se continua.

Paso 3. Seleccionar el recurso a mover. Se llamará X al Punto significativo con menor a (grado en que es cubierto por los recursos libres), es decir, se escoge el punto significativo menos cubierto. El recurso que se moverá será aquél cuya cobertura con respecto al punto X (Elemento de c_X de C de ese recurso) sea mayor.

Paso 4. Dirigir el recurso seleccionado hacia X y situarlo allí.

Paso 5. Volver al paso 1. (Se dejarán de mover recursos cuando no se satisfaga la condición del paso 2).

Situaciones especiales.

La situación de **interbloqueo** dentro de una zona en la que es necesario redistribuir se produce cuando cualquier movimiento de los recursos no provoca

un aumento lo suficientemente grande en la cobertura de los puntos significativos como para superar el umbral marcado. Esta situación es detectada en el algoritmo de distribución cuando dentro del conjunto de movimientos se produce un ciclo (o sea, una repetición de movimientos) sin ninguna modificación relevante del nivel de cobertura; si el algoritmo continuase en su ejecución, nunca terminaría por lo que una vez detectada se detiene su funcionamiento.

La detección de estas situaciones críticas no siempre es totalmente eficaz ya que la presencia de ciclos no tiene por que suponer siempre una situación donde no se puede alcanzar la cobertura mínima, lo que si suele significar es o bien una redistribución ineficiente ya que tiene que deshacer movimientos ya realizados o bien una redistribución de grandes movimientos de recursos, la cual, aún no siendo una situación de interbloqueo si da lugar a una situación crítica que debe tratarse como si lo fuera.

Una vez abortado el algoritmo se puede asumir alguna de las siguientes alternativas:

1. Tomar la cobertura mínima como umbral y mostar todos los movimientos menos el último (ya que éste culminaría el ciclo). Es una solución simple pero incompleta.
2. Reducir el umbral a un valor calculado a partir de los valores de cobertura que se hayan producido durante la ejecución del algoritmo. Después se ejecutaría de nuevo el algoritmo desde la situación inicial. El inconveniente de esta solución reside en que nuevas ejecuciones del algoritmo vuelvan a caer en interbloqueo y el proceso se ralentice.
3. Informar al usuario de la situación para que tome las medidas necesarias. Entre las acciones que se le pueden recomendar, puede estar la de solicitar la presencia de recursos que no están activos, etc. (No excluye las anteriores).

3.4. *Ejemplo.*

Dada la situación mostrada en la Fig. 42, donde los puntos nombrados con letras son Puntos significativos y los grises y negros dos tipos diferentes de recursos

libres, al aplicar el método descrito ocurre que los puntos cuyo **a** está por debajo del umbral son C, G e I. El primer recurso a mover es el negro situado en E, y tendrá que moverse hasta estar situado en C.

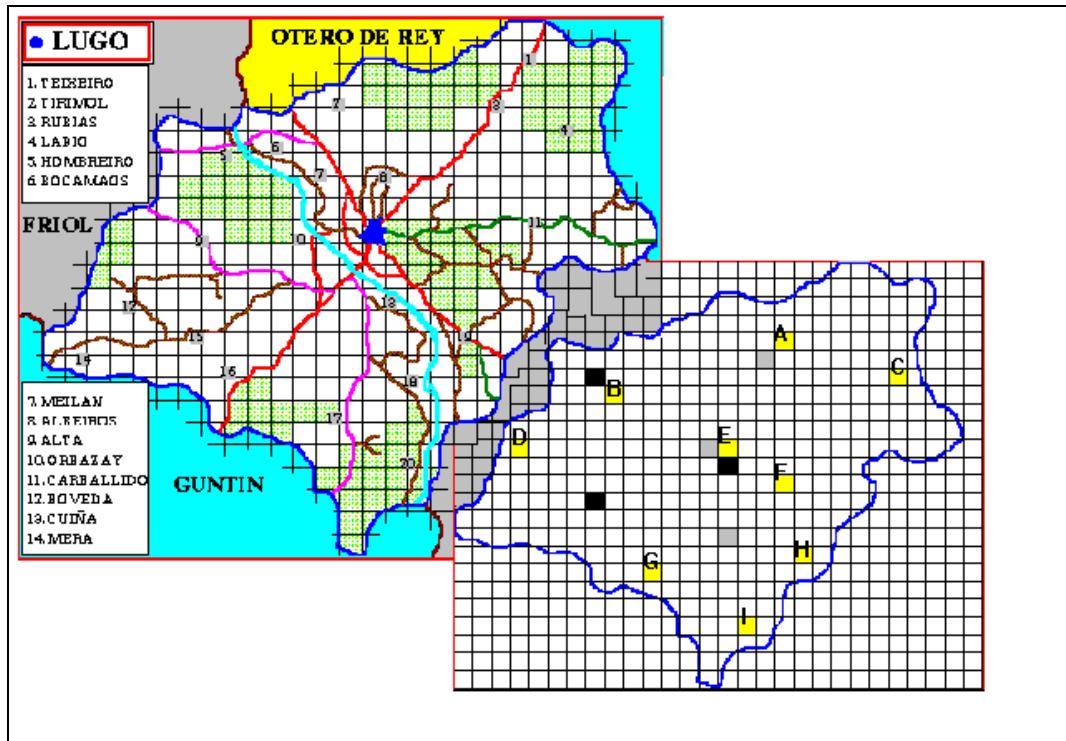


Figura 42: Ejemplo.

Al aplicar de nuevo el método lo que pasa es que hay que mover el recurso situado entre G y H hasta la posición de I. La tercera vez el consejo es mover el recurso situado entre D y G a la posición de G con lo que la zona queda mucho más equilibrada tal y como muestra la Fig. 43.

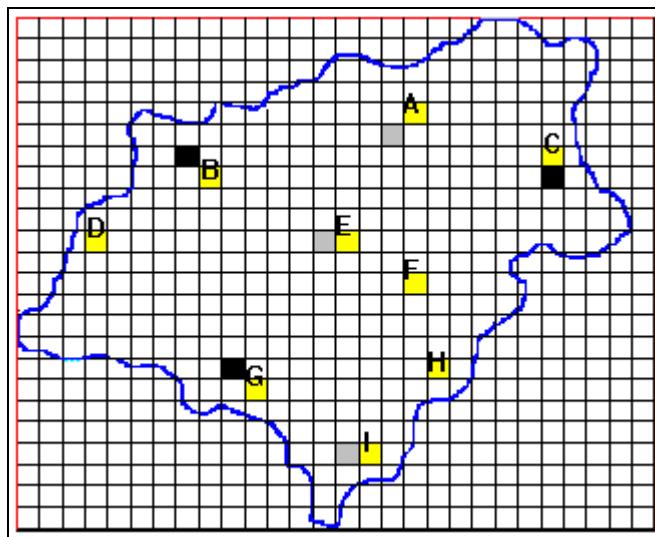


Figura 43: Resultado de la aplicación del algoritmo. Zona más equilibrada.

3.5. Implementación.

Paso 0: Tomar de la estructura interna los recursos no asignados que existan y de las Bases de Datos los referentes a los puntos significativos del municipio.

REPETIR Pasos 1-4 hasta salida del algoritmo

Paso 1: Determinar las áreas que no están suficientemente cubiertas por los recursos.

1.1 Calcular los datos sobre cobertura en los recursos libres.

1.1.1 Calcular la proximidad de los recursos a los puntos significativos.

Para Cada Recurso Hacer:

Para Cada Punto Significativo Hacer:

Calcular diferencia de coordenadas matriciales (dx y dy)

Calcular la distancia efectiva (distancia = ($dx > dy$)? $1/dx$: $1/dy$;

1.1.2 Calcular la cobertura de los recursos a los puntos significativos.

Para Cada Recurso i Hacer

Para Cada Punto Significativo P Hacer {

PROXIMIDAD = Recurso[i].Proximidad(P)

RELEVANCIA = RelevanciaPunto(P)

MOVILIDAD = Recurso[i].Movilidad (según su tipo)

COBERTURA =

Combinar(PROXIMIDAD,MOVILIDAD,RELEVANCIA) }

1.1.3 Calcular el valor de cobertura media de todos los recursos libres

Recurso[i].CoberturaMedia = Suma(CoberturaPuntos) / Nº de puntos significativos

1.2 Para todos los puntos significativos, determinar el grado en que es cubierto por los recursos libres

Para Cada Punto Significativo

COBERTURA = Suma(Cobertura de todos los REcursos) / Nº de recursos

Paso 2: Determinar si es necesario mover algún recurso. Para ello se determinan los puntos que tienen su cobertura por debajo del umbral y se destaca al que tenga mínima cobertura. Si no existe ningún recurso que reúna estas condiciones entonces la zona está cubierta por encima del valor umbral y no es necesario redistribuir

2.1 Si hay algún punto descubierto

Obtener el Indice del recurso que cubre mejor al punto descubierto

```
int dxPuntoMin = Puntos->IndexMenosCubierto();
```

SI NO Salir del Algoritmo (ir A Paso 5)

```
if (dxPuntoMin == -1) {
```

```
    MessageBox("No es necesario redistribuir");
```

```
    break;
```

```
}
```

Paso 3: Seleccionar el recurso a mover. Se llamará X al Punto Significativo con menor α (grado en que es cubierto por los recursos libres, es decir, se escoge el

punto significativo menos cubierto. El recurso que se moverá será aquel cuya cobertura con respecto al punto X (elemento Cx de C de ese recurso) sea mayor.

3.1 Obtener el recurso que cubre mejor al punto descubierto y que no esté ya localizado en double cobmax = 0;

```
int dxRecursoMaxCob = -1;

int ii;
for (ii = 0; ii < MAXRECURSOS; ii++) {

    double cobpunto = Recursos[ii]->ValorPointCobertura(dxPunto);

    double proxpunto = Recursos[ii]-
        >ValorPointProximidad(dxPunto);

    // se comprueban los recursos que estén a un mínimo de distancia

    if (cobpunto >= cobmax      && proxpunto < MINDISTANCIA) {

        cobmax = Recursos[ii]->ValorPointCobertura(dxPunto);

        dxRecursoMaxCob = ii;

    }

}
```

Paso 4: Dirigir el recurso seleccionado hacia X y situarlo allí

4.1 Representar por pantalla el movimiento a realizar

4.2 Representar en el mapa el movimiento

4.3 Modificar la posición del recurso

Paso 5: Fin de la redistribución de recursos

Método para detectar el interbloqueo:

Introducir cada vez que se genere un nuevo movimiento de recursos el siguiente código:

Insertar(MovimientoNuevo, ListaMovimientos)

Si (Hay_Ciclo(ListaMovimientos) = VERDAD) entonces ACCIONES
INTERBLOQUEO

sino CONTINUAR CON EL ALGORITMO

3.6. Complejidad del algoritmo de redistribución de recursos.

Los parámetros determinantes de la complejidad de este algoritmo van a ser:

R: número de recursos libres que van a ser tratados en el algoritmo.

P: número de puntos significativos representados en la zona.

N: número de iteraciones ya realizadas.

c: constante multiplicativa que indica el coste de las operaciones.

Complejidad del cuerpo del bucle del algoritmo:

Paso 1:

Paso 1.1.1 Calculo de la proximidad ($c \cdot R \cdot P$).

Paso 1.1.2 Calculo de la cobertura ($c \cdot R \cdot P$).

Paso 1.1.3. Calculo de la cobertura media ($c \cdot R \cdot P$).

Paso 1.1.4. Calculo de la cobertura de los puntos ($c \cdot R \cdot P$)

Paso 2:

Paso 2.1: Determinar el punto con menos cobertura y debajo del umbral (P).

Paso 2.2 Si no hay ninguno CASO MEJOR.

Paso 3: Seleccionar el recurso a mover (R).

Paso 4: Detectar interbloqueos (implementación actual c, en implementación completa N).

Casos que se pueden presentar:

Caso mejor: Sólo una ejecución del algoritmo

$$O(R, P) = \text{Max}((c \cdot R \cdot P), (c \cdot R \cdot P), P, R, N) = c \cdot R \cdot P.$$

el algoritmo es de una complejidad a lo sumo cuadrática al número de recursos y a los puntos significativos.

Casos medio y peor: La complejidad de los casos peor y medio depende de muchos factores lo cual hace imposible su determinación a priori. Factores influyentes son: el nº de recursos, el nº de puntos significativos de la comarca, la relevancia de éstos, la ponderación de los factores, la distancia entre puntos significativos y recursos... Se podrían distinguir los siguientes casos:

Si hay más o menos proporción entre los recursos y los puntos significativos con un umbral de cobertura razonable se podría pensar en un caso medio en el cual el número de iteraciones sería N , que dependería de la distancia de los recursos a los puntos. Por lo tanto la complejidad de este caso medio sería $N \cdot R \cdot P$.

En el caso de que hubiera desproporción entre recursos y puntos, muchos recursos y pocos puntos darían una complejidad de $N \cdot R \cdot P$ ($N < R, P$) en el caso general o de $R \cdot P$ en el mejor de los casos. Si existen muchos puntos significativos y pocos recursos se daría una situación de interbloqueo que sería detectada muy pronto, o sea, una complejidad de $R \cdot P$. En ningún caso el algoritmo entraría en un bucle infinito o llegaría a una complejidad NP .

3.7. *Interfaces.*

Los interfaces de que dispone el usuario para esta aplicación son uno de inicio (fig. 44) y uno de resultados (fig. 45):

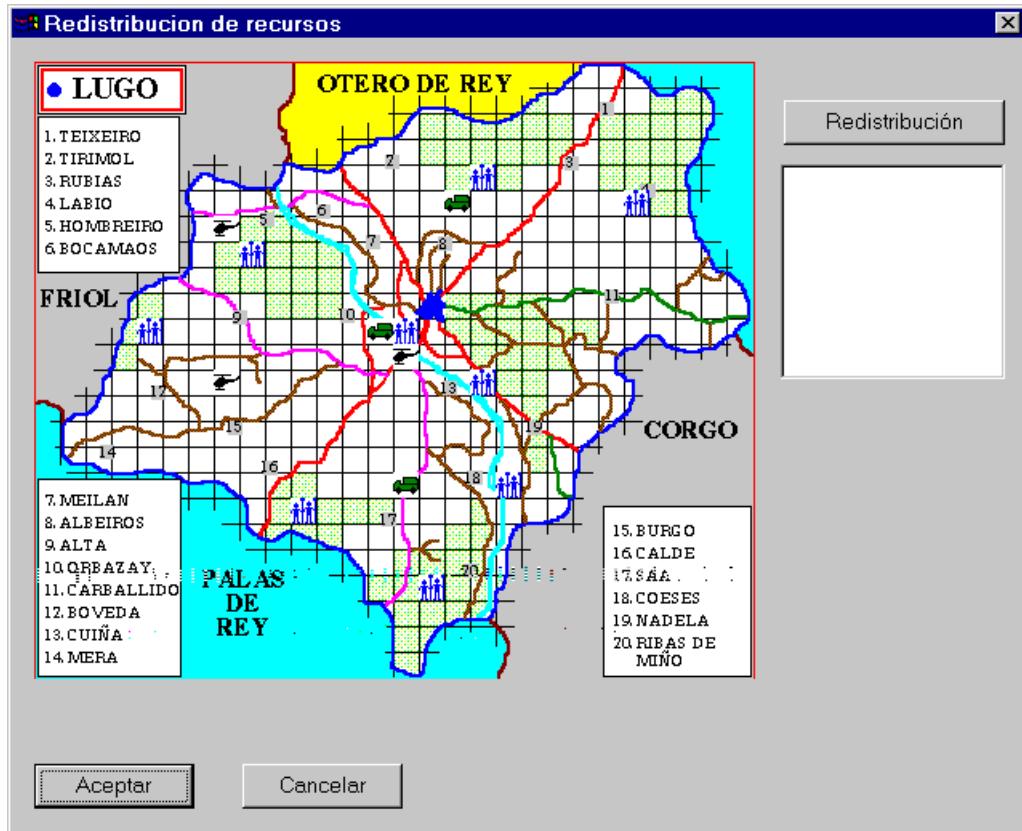


Figura 44: Interfaz de inicio de la Redistribución de recursos.

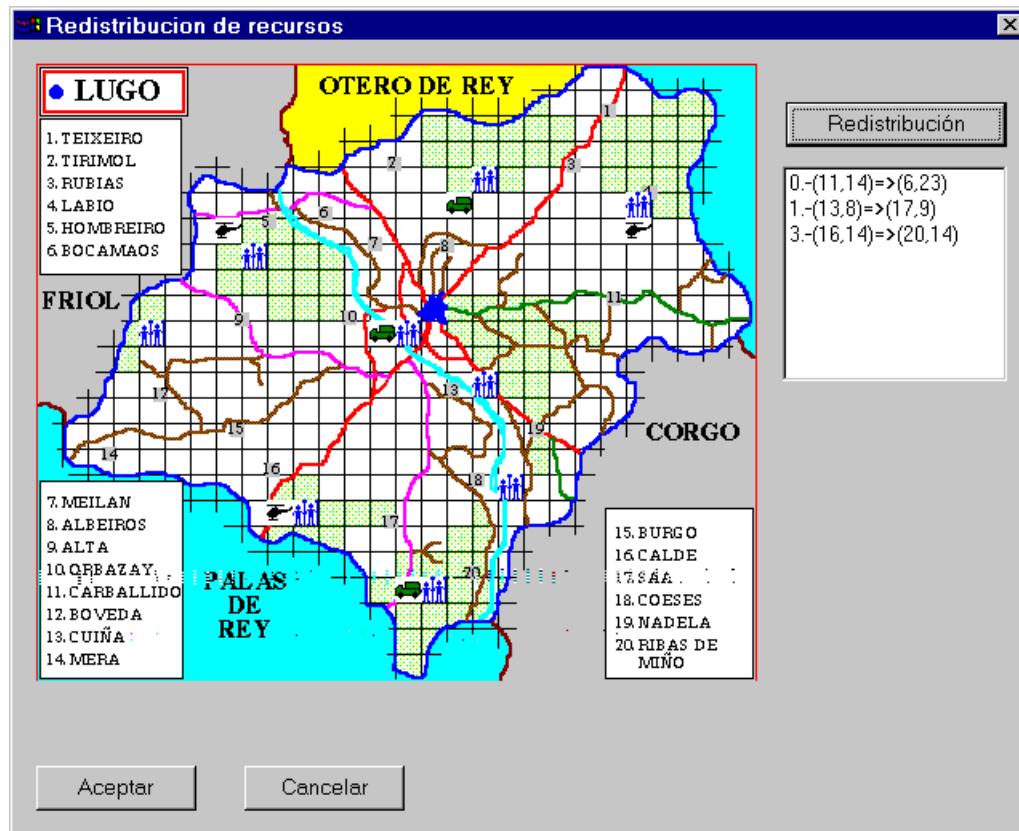


Figura 45: Interfaz resultado de la Redistribución de recursos

Capítulo 5

Validación y Verificación del Sistema.

Aunque es imprescindible una evaluación pormenorizada del sistema, no es el objetivo de esta memoria detallar todas las operaciones llevadas a cabo para tal fin. En este capítulo se muestra una idea general de cómo se han llevado a cabo los procesos de verificación y validación de INCEND-IA, viendo con detalle las respuestas del sistema en un caso real y haciendo una aproximación al porcentaje de éxito de las respuestas del sistema.

1. Evaluación del Sistema.

La evaluación de un Sistema Basado en el Conocimiento no es una tarea sencilla, debido, entre otras, a las siguientes razones:

- Las expectativas con respecto a sus prestaciones son frecuentemente vagas y confusas.
- Muchas veces los casos típicos que debe resolver el sistema son los mismos que se han usado para la adquisición de conocimiento.
- No es posible en un caso real disponer de un *juego de pruebas* que pueda asegurar que el comportamiento del sistema es el ideal.

Además, comparándolo con la evaluación de un software convencional, la evaluación de un SBC posee las siguientes características diferenciales:

- Los SBC no son por naturaleza objetivos.

- Los SBC gestionan habitualmente cierta cantidad de incertidumbre.
- En el software convencional puede usarse el juego de ensayo.
- En el software convencional normalmente no se cuestiona la corrección de los resultados.

Se muestra a continuación en qué consiste el proceso de evaluación en general [Gómez, 97]:

Quién hace la acción	Acciones	Qué usa la acción	Quién genera la entrada		
Evaluador	Planteamiento de la evaluación	Información proceso de construcción	Constructor		
	Objetivo				
	Criterios				
	Técnica de valoración				
	Carga de prueba				
	Patrón de referencia				
Evaluador	Realización de la evaluación	Sistema	Constructor		
	Examen o estudio del sistema				
	Juicio de valor del aspecto a				
↓					
Resultado de la evaluación: R					
Evaluador	Comparación del resultado con el patrón de referencia.	Patrón de referencia: P	Constructor		
	Toma de Decisión		Evaluador		
	R ≠ P Volver a atrás: Modificar el sistema.	Objetivo de la evaluación	Constructor		
	R = P Seguir adelante con la construcción del sistema				

Tabla 71: Proceso de evaluación.

2. Verificación del Sistema.

La verificación del sistema consiste en comprobar la corrección de los modelos conceptuales, formales y computables del sistema. La debe realizar un Ingeniero de Conocimiento distinto al que desarrolló el sistema. Lo que se verificó en INCEND-IA a lo largo de todo el proceso de creación, básicamente ha sido:

Redundancias:

- Identidad entre elementos.
- Elementos sin conocimiento.
- Circularidades.
- Partes innecesarias.

Compleitud:

- Elementos ausentes.
- Elementos inalcanzables.

Consistencia: (Depende de cada formalismo de representación utilizado, de los tipos de lógicas y de los mecanismos de gestión de incertidumbre):

- Elementos contradictorios.
- Incompatibilidades con las restricciones.

3. Validación del Sistema.

La validación consiste en comprobar que el sistema responde adecuadamente a los fines para que fue diseñado. La debe realizar el experto, o utilizarse conjuntos de casos de prueba (existen indicaciones sobre el número de casos de prueba que deben usarse [Caraça-Valente, 99]), pero en general, se deben realizar los siguientes procesos [Gómez, 97], tabla 72:

		Carga de Prueba	Patrón de referencia
Comparación frente a resultados conocidos		Casos históricos con solución conocida	Comparar resultado del sistema con el conocido
Comparación frente a modelos matemáticos	Modelo Teórico	Casos aleatorios	Comparar resultado del sistema con la solución del modelo para el caso
	Modelo lineal		
Comparación frente a un experto	Juego aleatorio	Casos aleatorios	Comparar resultado del sistema con el del experto
	Juego de ensayo	Casos seleccionados	
	Ensayo en paralelo	Casos reales	

Análisis de sensibilidad	Sensibilidad a variaciones en las entradas	Casos aleatorios donde se provocan variaciones en la entrada	Las salidas de los casos variados deben ser semejantes
	Sensibilidad a errores en las entradas	Casos con entradas inválidas	No debe producirse salida y debe dar mensaje de error

Tabla 72: Proceso de validación.

También deben ser evaluadas la usabilidad y la utilidad del sistema, cuando el sistema está funcionando en condiciones reales.

3.1. *Validación del sistema INCEND-IA.*

Para validar el sistema INCEND-IA se han usado todos los mecanismos de validación descritos, salvo el de comparación frente a modelos matemáticos. A continuación se muestra uno (de los varios realizados), proceso completo de comparación frente a resultados conocidos, la finalización del ejemplo que se ha venido mostrando hasta el momento, y la evaluación del resultado.

El ciclo real que se utiliza en esta validación es el siguiente:

Día	NumIncendios	Especial	Personas	Vehículos
06-agosto-98	1	1	1	0
07-agosto-98	1	1	2	1
08-agosto-98	5	3	8	6
09-agosto-98	4	6	8	4
10-agosto-98	8	9	14	11
11-agosto-98	16	27	25	39
12-agosto-98	7	16	18	12
13-agosto-98	6	12	14	8
14-agosto-98	7	11	10	7
15-agosto-98	8	8	8	6
16-agosto-98	9	18	26	23
17-agosto-98	3	4	6	4
18-agosto-98	4	8	4	4

Tabla 73: Ciclo real usado para la validación (1998).

Factores de ciclo:

El usuario el primer día del ciclo introduce los siguientes valores: {Sequedad: Media/0.45, LLuvia: Tormenta/0.5, Influencia social: Mucha/0.85}. Lo cual resulta $M = 0.6$ (fig. 47).

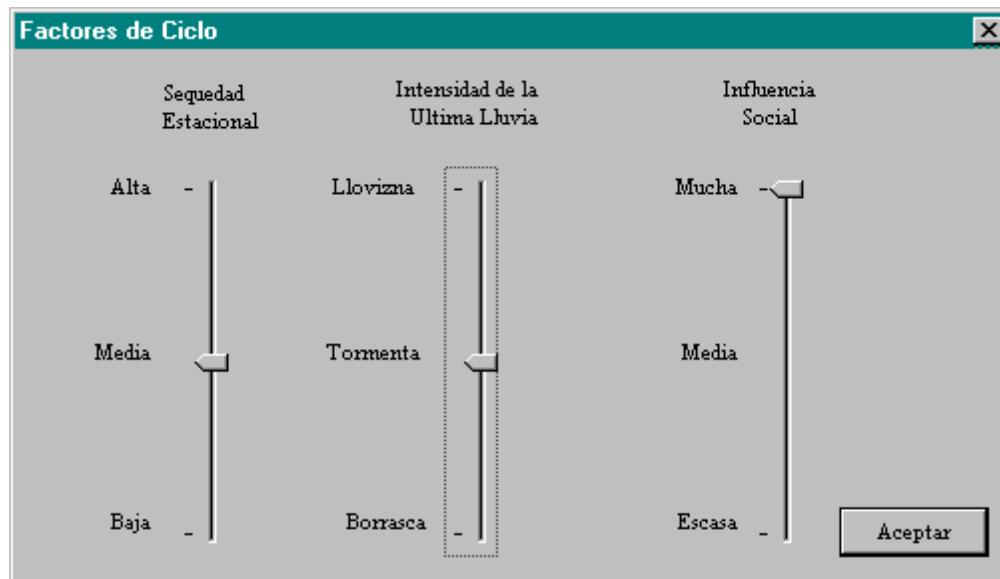


Figura 46: Factores de ciclo.

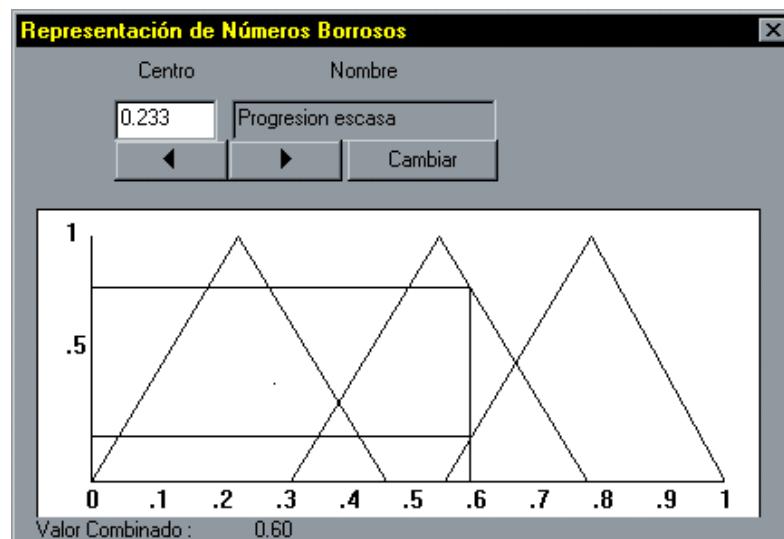


Figura 47: Pertenencia a las Categorías Prototípicas Deformables Borrosas.

Con $\mu_{p_2} = 0.82$ y $\mu_{p_3} = 0.28$, la deformación de las Categorías Prototípicas Deformables Borrosas afines hacia el ciclo real es la siguiente:

$$C_{real}(v_1 \dots v_n) = 0.82 * \begin{pmatrix} 6 & 2 & 6 \\ 3 & 4 & 6 \\ 1 & 1 & 1 \\ 8 & 11 & 19 \\ 4 & 6 & 11 \\ 5 & 6 & 12 \\ 4 & 5 & 11 \end{pmatrix} + 0.28 * \begin{pmatrix} 8 & 13 & 15 \\ 9 & 12 & 19 \\ 1 & 1 & 1 \\ 31 & 53 & 85 \\ 8 & 16 & 17 \\ 13 & 21 & 33 \\ 5 & 13 & 18 \end{pmatrix}$$

Y la definición del Ciclo real actual queda de la siguiente forma:

	Sector 1	Sector 2	Sector 3
Media de días:	7	5	9
Media de incendios/día:	5	7	10
Mínimo de inc/día:	1	1	1
Máximo de inc/día:	15	24	39
Nº especialistas/día:	6	9	14
Nº de brigadas/día:	8	11	19
Nº de medios/día:	5	8	14

Tabla 74: Definición del ciclo real para $\mu p_2 = 0.82$ y $\mu p_3 = 0.28$.

Una vez introducidos los factores de ciclo, la predicción inicial comenzando por el 6 de Agosto, queda como sigue (fig. 48). En un cuadro de texto en la parte superior se muestra una descripción genérica del sector en el que se encuentra la predicción, que también es útil para el usuario:

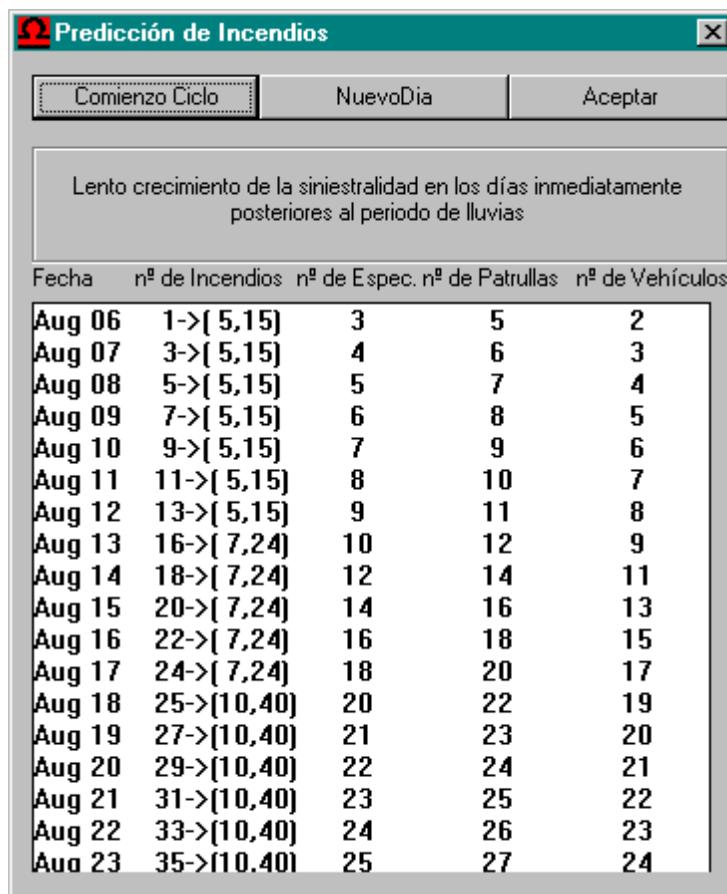


Figura 48: Predicción al comienzo del nuevo ciclo.

Lo que realmente ocurrió frente a lo que se ha predecido para el día 6 de Agosto es lo siguiente:

Día	NumIncendios	Especial	Personas	Vehículos
06-agosto-98	1	1	1	0
Predicción	1	3	5	2

El día 7 por la mañana el usuario introdujo los siguientes factores diarios:

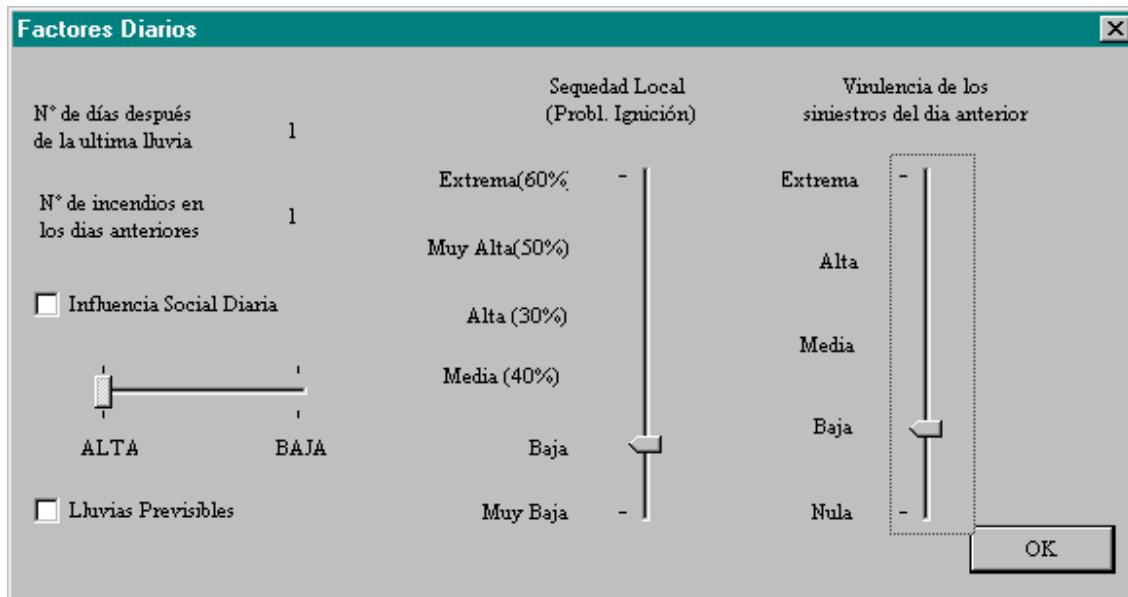


Figura 49: Factores diarios del 7 de Agosto.

El sistema generó la siguiente predicción:

Predicción de Incendios				
Comienzo Ciclo		NuevoDia		Aceptar
Lento crecimiento de la siniestralidad en los días inmediatamente posteriores al periodo de lluvias				
Fecha	nº de Incendios	nº de Espec.	nº de Patrullas	nº de Vehículos
Aug 07	3->[5,15]	4	6	3
Aug 08	5->[5,15]	5	7	4
Aug 09	7->[5,15]	6	8	5
Aug 10	9->[5,15]	7	9	6
Aug 11	11->[5,15]	8	10	7
Aug 12	13->[5,15]	9	11	8
Aug 13	16->[7,24]	10	12	9
Aug 14	18->[7,24]	12	14	11
Aug 15	20->[7,24]	14	16	13
Aug 16	22->[7,24]	16	18	15
Aug 17	24->[7,24]	18	20	17
Aug 18	25->[10,40]	20	22	19
Aug 19	27->[10,40]	21	23	20
Aug 20	29->[10,40]	22	24	21
Aug 21	31->[10,40]	23	25	22
Aug 22	33->[10,40]	24	26	23
Aug 23	35->[10,40]	25	27	24
Aug 24	37->[10,40]	26	28	25

Figura 50: Predicción del 7 de Agosto.

Ahora se puede analizar lo ocurrido frente a lo predecido:

Día	NumIncendios	Especial	Personas	Vehículos
07-ago-98	1	1	2	1
Predicción	3	4	6	3

Los factores diarios del día 8 fueron los siguientes:

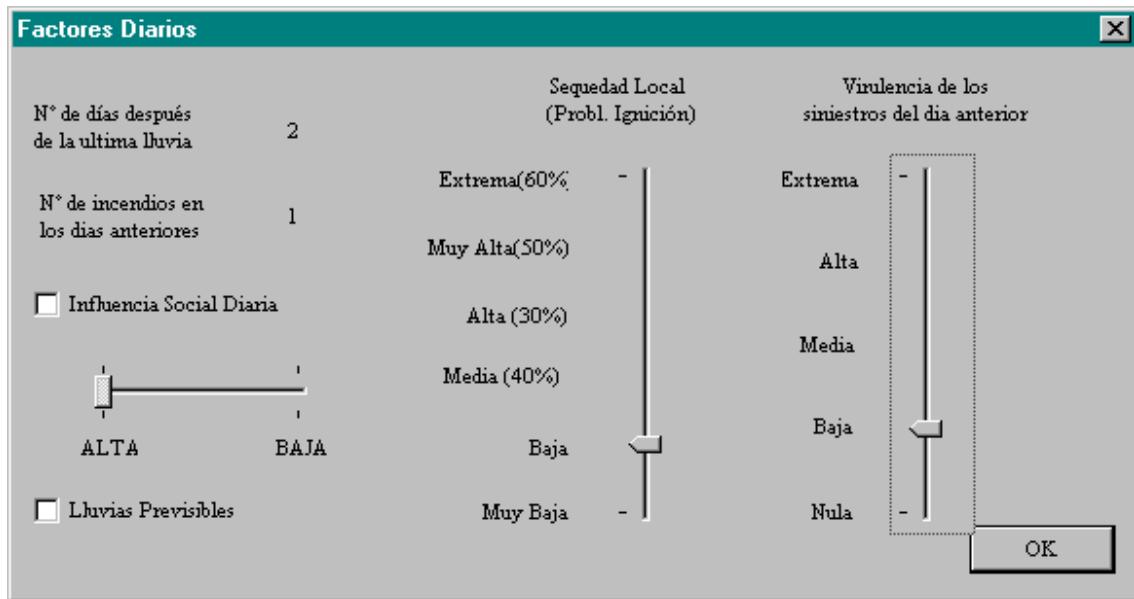


Figura 51: Factores diarios del 8 de Agosto.

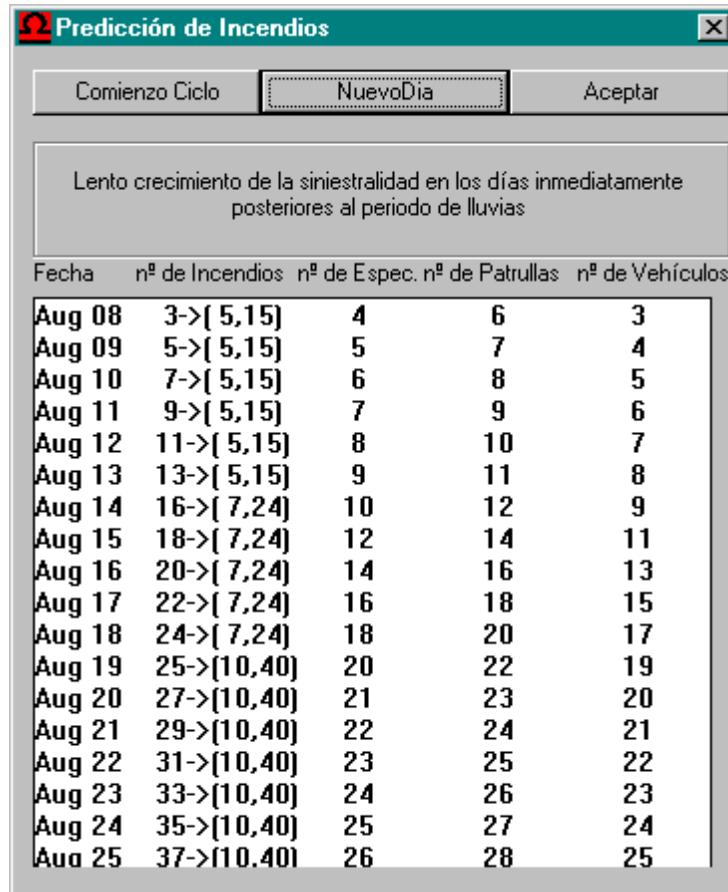


Figura 52: Predicción del 8 de Agosto.

Ahora se puede analizar lo ocurrido frente a lo predecido, como se puede apreciar se ha estancado la predicción:

Día	NumIncendios	Especial	Personas	Vehículos
08-ago-98	5	3	8	6
Predicción	3	4	6	3

El día 9 el ususario ha introducido los siguientes factores, lo que ha provocado un cambio de sector.

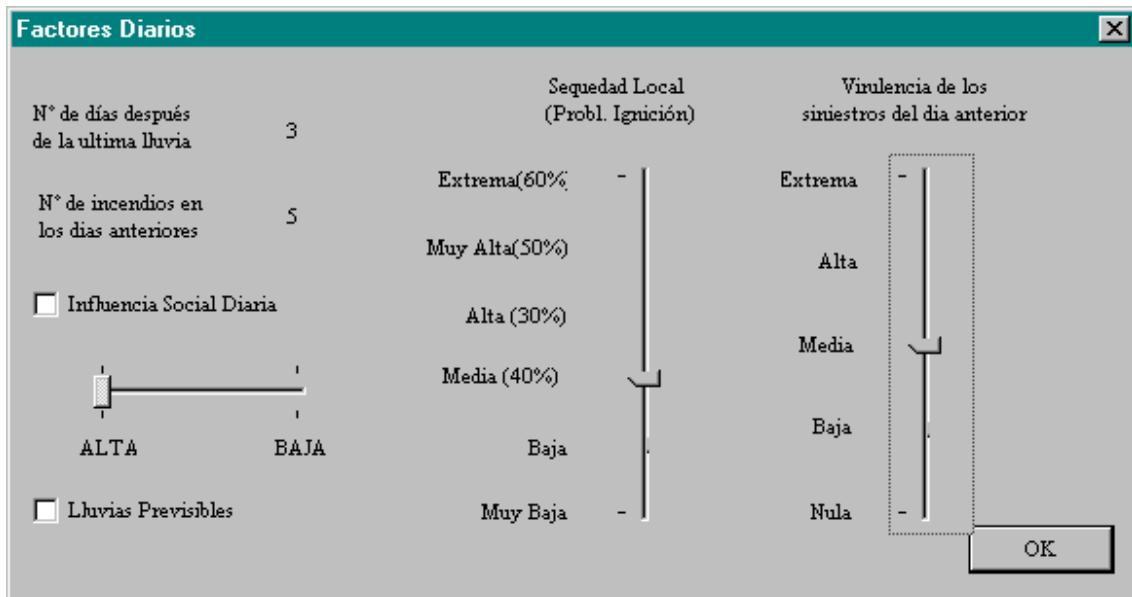


Figura 53: Factores diarios del 9 de Agosto.

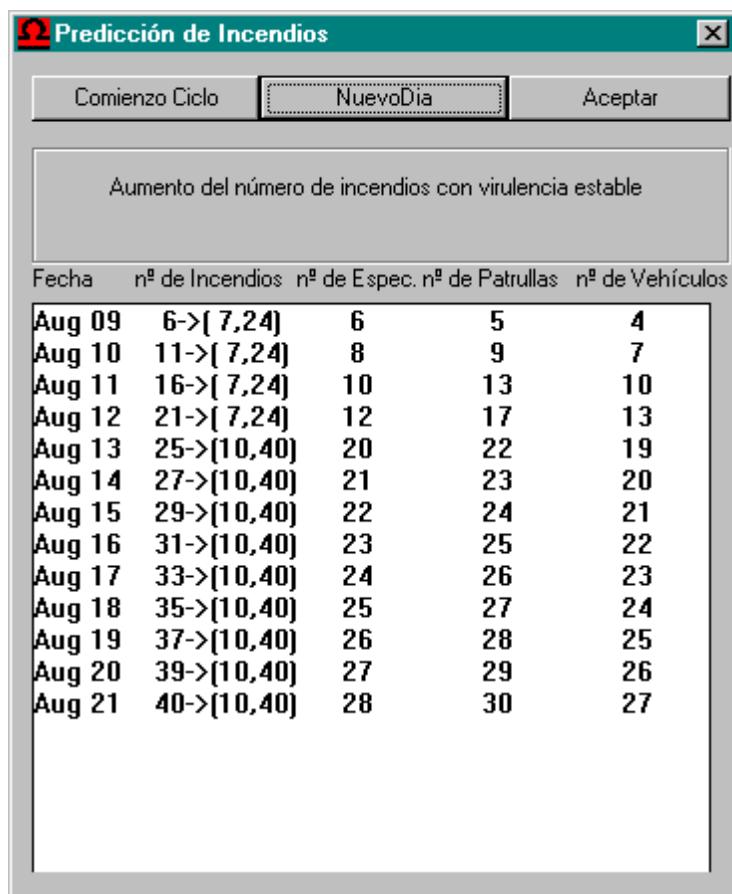


Figura 54: Predicción del 9 de Agosto.

Se puede observar que ha cambiado el mensaje de información del sector y la predicción se ha desplazado hacia el segundo sector. Comparativamente:

Día	NumIncendios	Especial	Personas	Vehículos
09-agosto-98	4	6	8	4
Predicción	6	6	5	4

El día 10 el usuario introdujo los mismos factores diarios, la predicción se estanca.

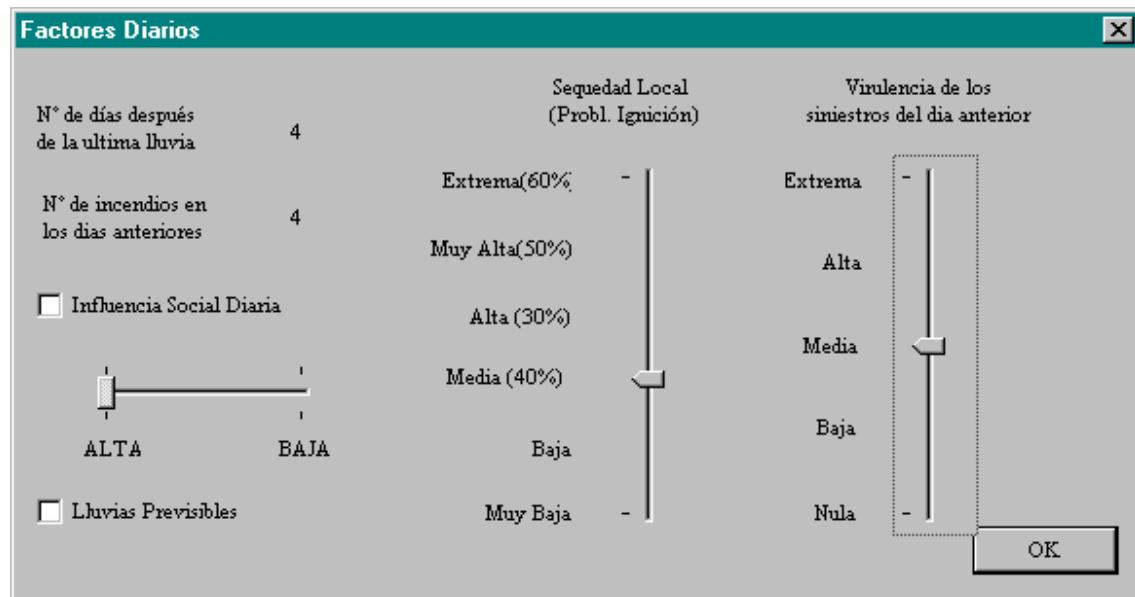


Figura 55: Factores diarios del 10 de Agosto.

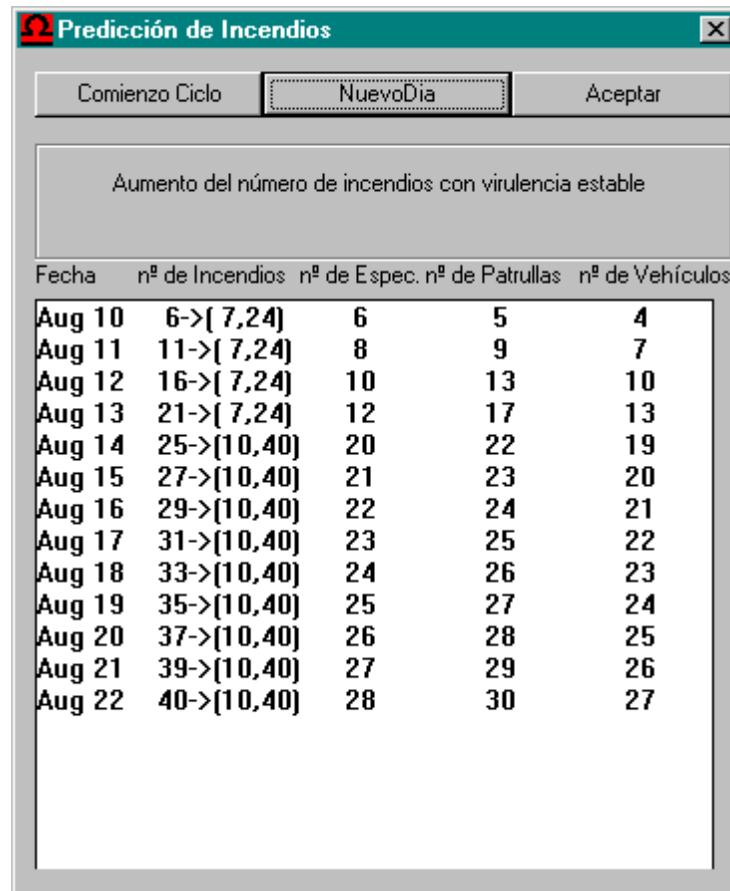


Figura 56: Predicción del 10 de Agosto.

Comparando lo ocurrido con lo que se predijo:

Día	NumIncendios	Especial	Personas	Vehículos
10-agosto-98	8	9	14	11
Predicción	6	6	5	4

El día 11 el usuario introdujo sus factores diarios e incluyó el de *influencia social baja*, por lo que se le mostró un mensaje de aviso y se continuó la predicción.

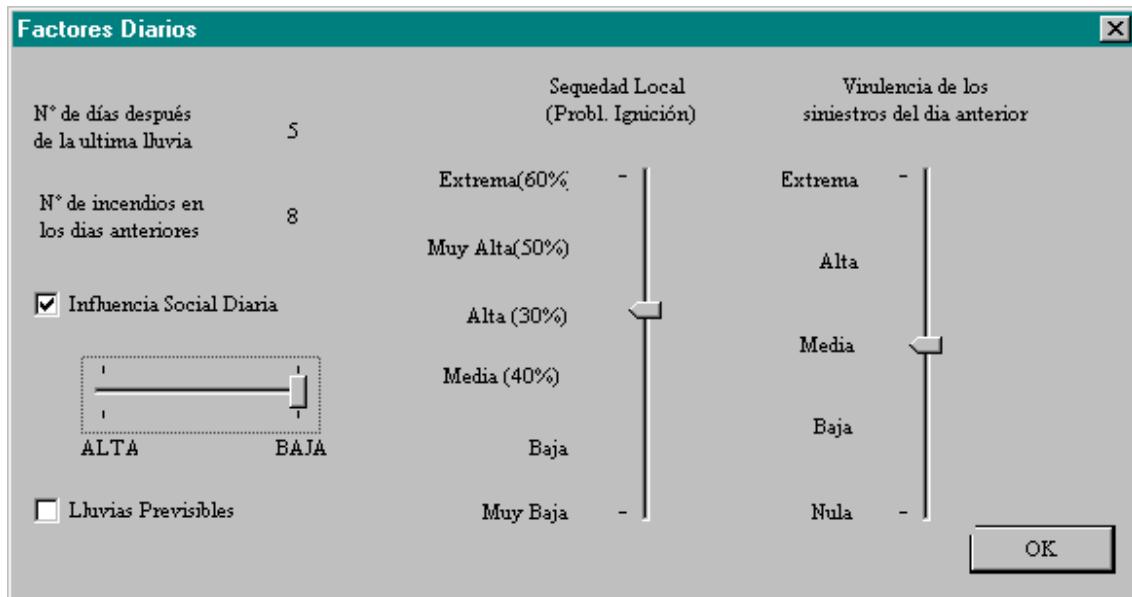


Figura 57: Factores diarios del 11 de Agosto.

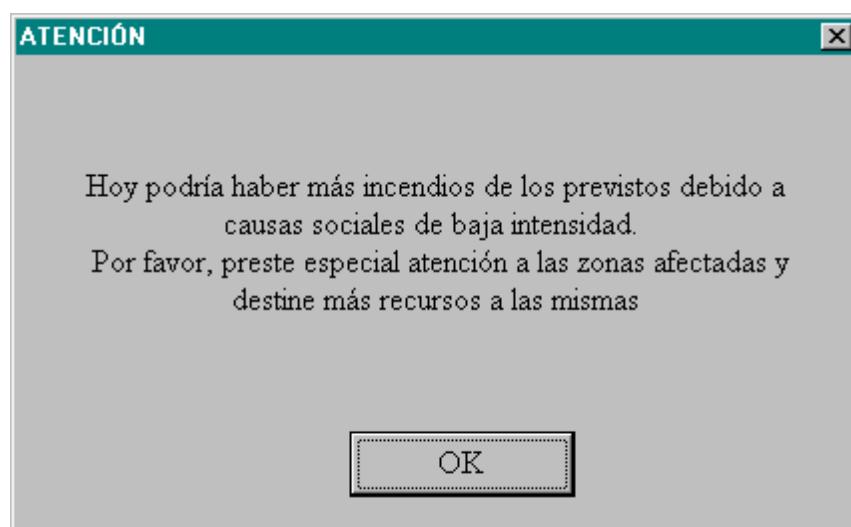


Figura 58: Mensaje de *Influencia Social Baja*.

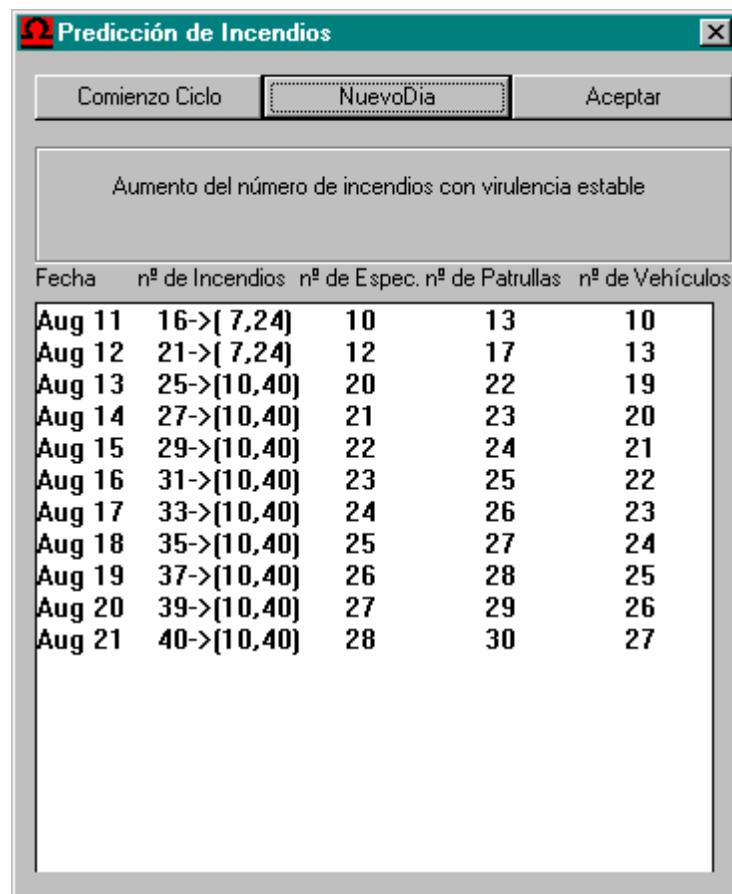


Figura 59: Predicción del 11 de Agosto.

Comparando lo ocurrido con lo que se predijo:

Día	NumIncendios	Especial	Personas	Vehículos
11-agosto-98	16	27	25	39
Predicción	16	10	13	10

El número de medios que se ha tenido que utilizar el día 11 provoca un cambio de sector en la predicción. Al comienzo del día 12 el usuario introdujo los siguientes factores diarios:

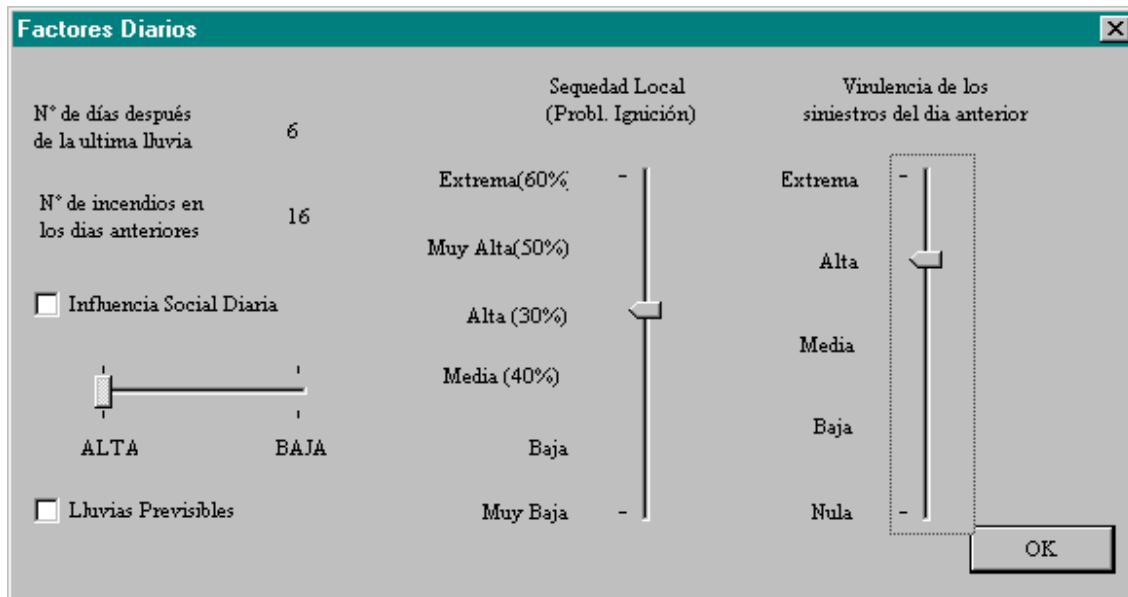


Figura 60: Factores diarios del 12 de Agosto.

Y el resultado de la predicción fue el siguiente:

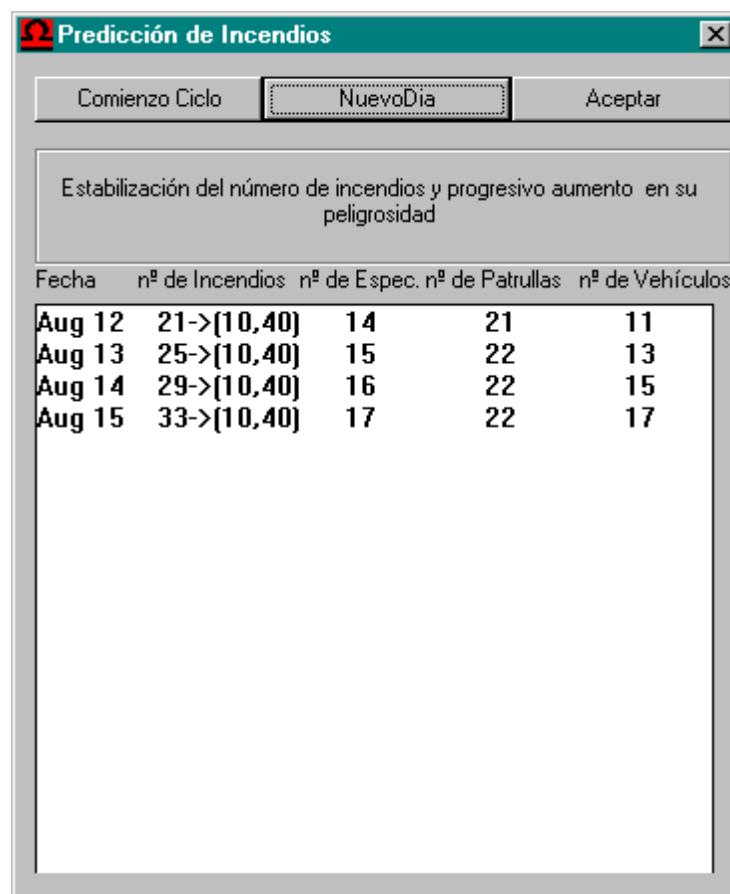


Figura 61: Predicción del 12 de Agosto.

Comparando lo ocurrido con lo que se predijo:

Día	NumIncendios	Especial	Personas	Vehículos
12-agosto-98	7	16	18	12
Predicción	21	14	21	11

Los datos que el usuario introdujo el día 13 fueron:

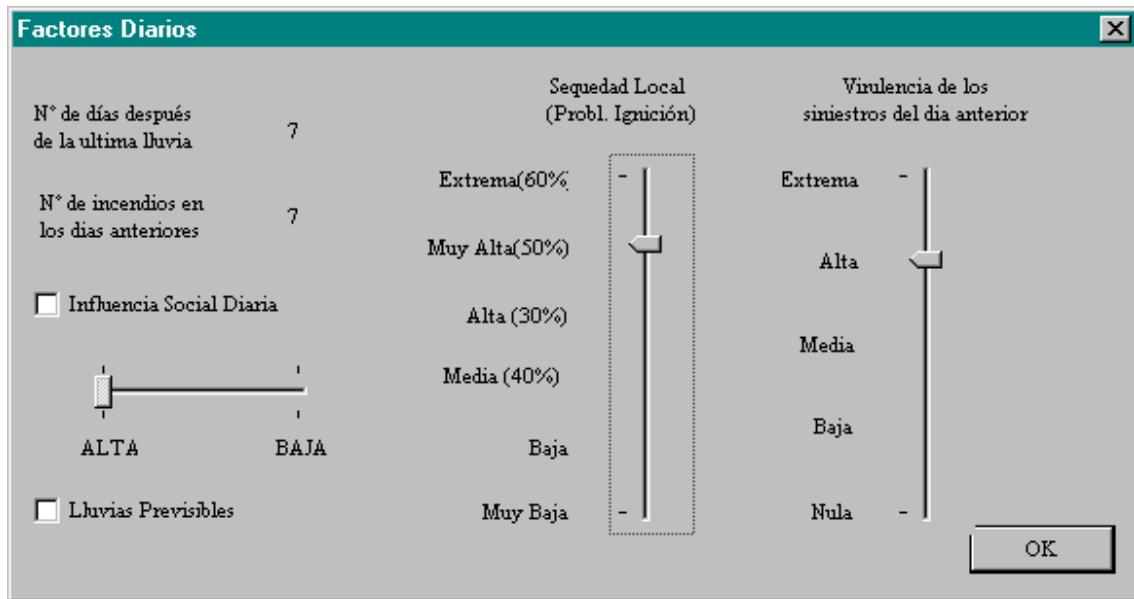


Figura 62: Factores diarios del 13 de Agosto.

Y la predicción la que sigue:

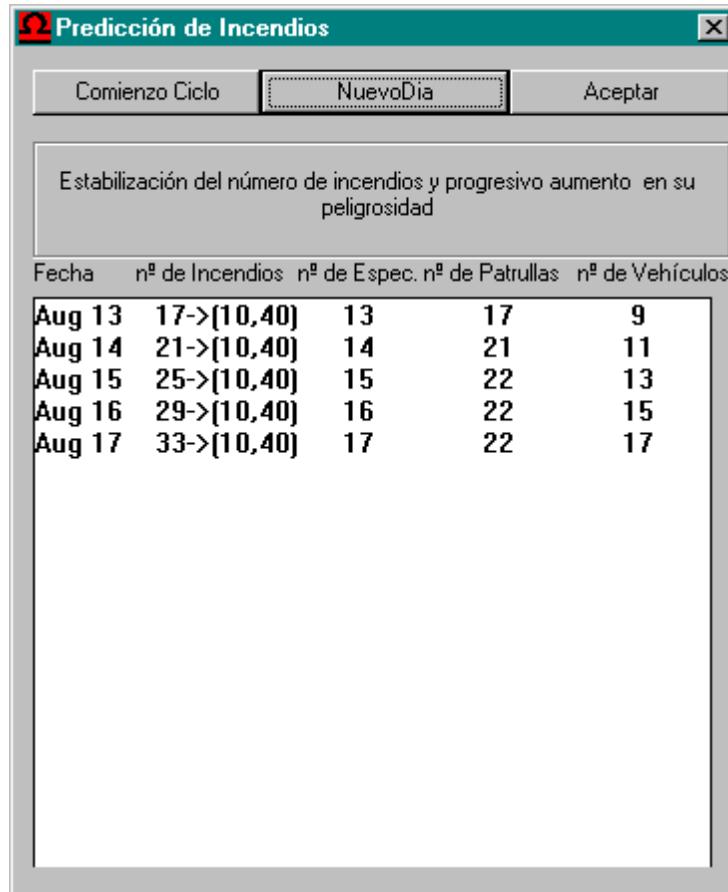


Figura 63: Predicción del 13 de Agosto.

Comparando lo ocurrido con lo que se predijo:

Día	NumIncendios	Especial	Personas	Vehículos
13-agosto-98	6	12	14	8
Predicción	17	13	17	9

Los datos que el usuario introdujo el día 14 fueron:

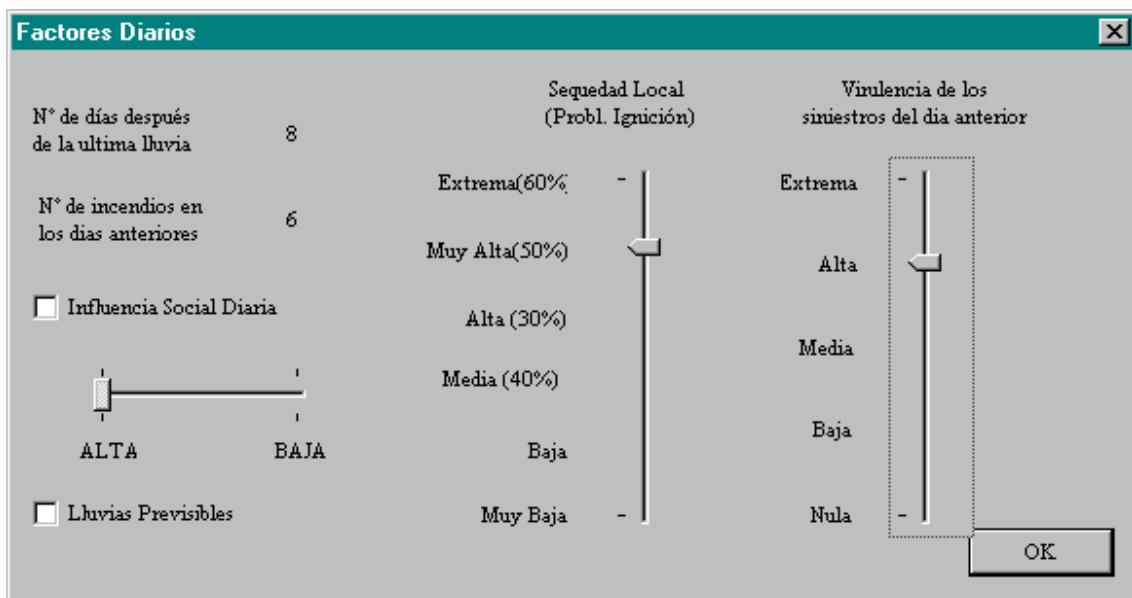


Figura 64: Factores diarios del 14 de Agosto.

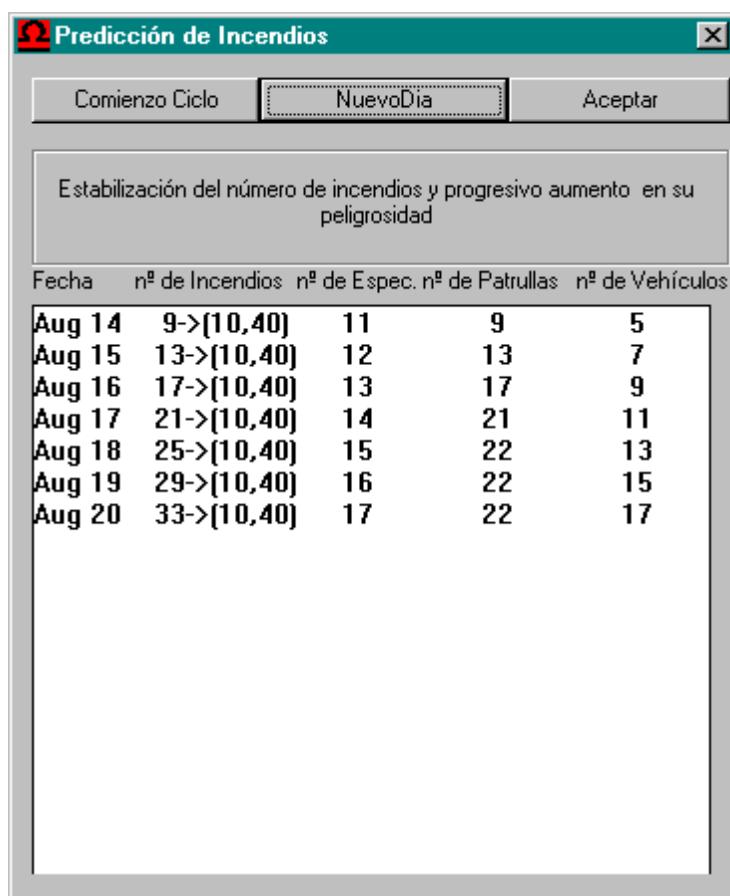


Figura 65: Predicción del 14 de Agosto.

Comparando lo ocurrido con lo que se predijo:

Día	NumIncendios	Especial	Personas	Vehículos
14-ago-98	7	11	10	7
Predicción	9	11	9	5

Los datos que el usuario introdujo el día 15 fueron:

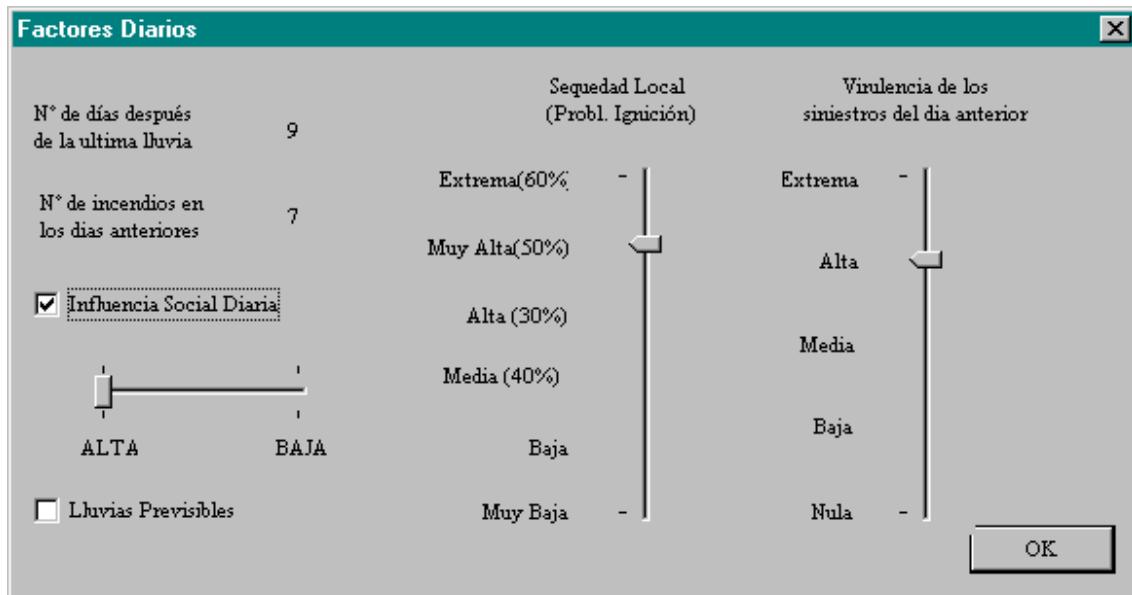


Figura 66: Factores diarios del 15 de Agosto.

Obsérvese que se estanca la predicción, pero el usuario recibe el siguiente aviso:

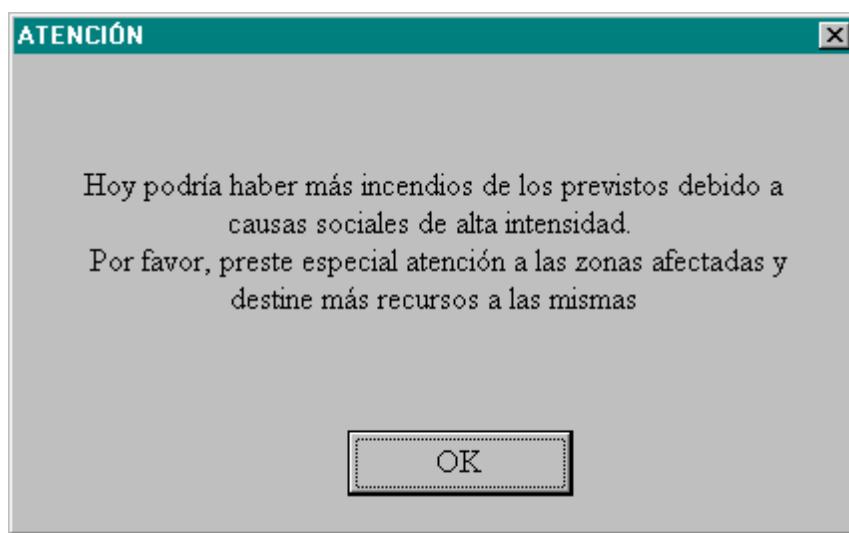


Figura 67: Mensaje de *Influencia Social Alta*.

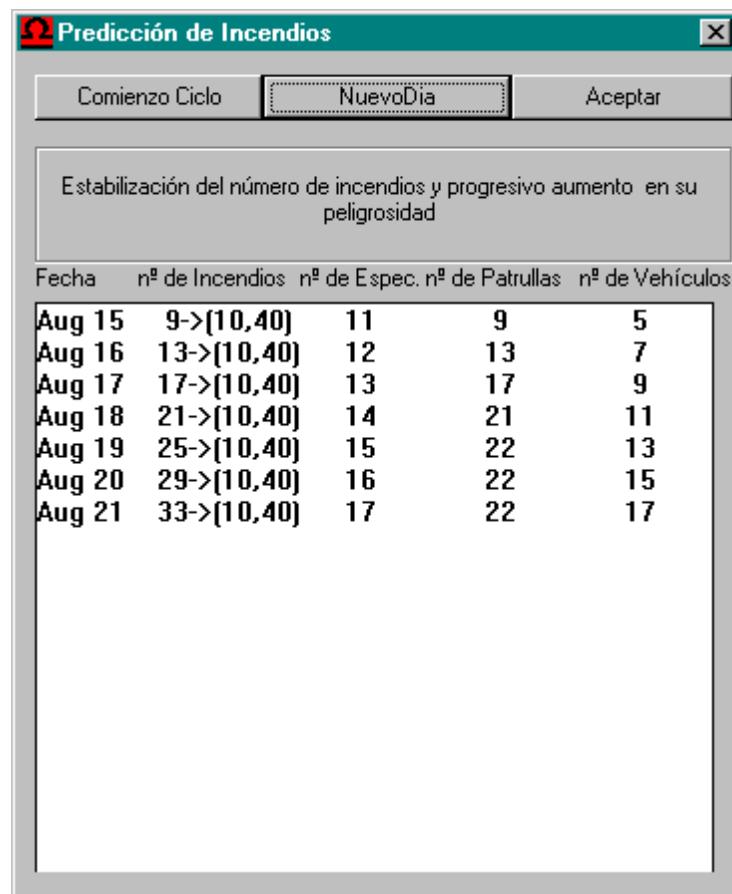


Figura 68: Predicción del 15 de Agosto.

Comparando lo ocurrido con lo que se predijo:

Día	NumIncendios	Especial	Personas	Vehículos
15-agosto-98	8	8	8	6
Predicción	9	11	9	5

El día 15 coincidió en sábado, festivo en muchos pueblos en Galicia y con gran índice de actividad a partir de la finalización de las verbenas populares. Por ello, se esperan especialmente complicadas las primeras horas del domingo 16 y todo este día. Los datos que el usuario introdujo el día 16 fueron:

Factores Diarios

Nº de días después de la última lluvia	10	Sequedad Local (Probl. Ignición)	Extrema(60%)	Virulencia de los siniestros del dia anterior
Nº de incendios en los días anteriores	9	Muy Alta(50%)	[Slider]	Extrema
<input type="checkbox"/> Influencia Social Diaria		Alta (30%)	[Slider]	Alta
		Media (40%)	[Slider]	Media
		Baja	[Slider]	Baja
<input type="checkbox"/> Lluvias Previsibles		Muy Baja	[Slider]	Nula

ALTA BAJA

OK

Figura 69: Factores diarios del 16 de Agosto.

Predicción de Incendios

Comienzo Ciclo	<input type="button" value="NuevoDia"/>	Aceptar		
Estabilización del número de incendios y progresivo aumento en su peligrosidad				
Fecha	nº de Incendios	nº de Espec.	nº de Patrullas	nº de Vehículos
Aug 16	9->[10,40]	11	9	5
Aug 17	13->[10,40]	12	13	7
Aug 18	17->[10,40]	13	17	9
Aug 19	21->[10,40]	14	21	11
Aug 20	25->[10,40]	15	22	13
Aug 21	29->[10,40]	16	22	15
Aug 22	33->[10,40]	17	22	17

Figura 70: Predicción del 16 de Agosto.

Comparando lo ocurrido con lo que se predijo:

Día	NumIncendios	Especial	Personas	Vehículos
16-agosto-98	9	18	26	23
Predicción	9	11	9	5

Los datos que el usuario introdujo el día 17 fueron los siguientes. Nótese que hay una previsión de lluvias dada por los servicios meteorológicos, por lo que el usuario recibe un mensaje de aviso.

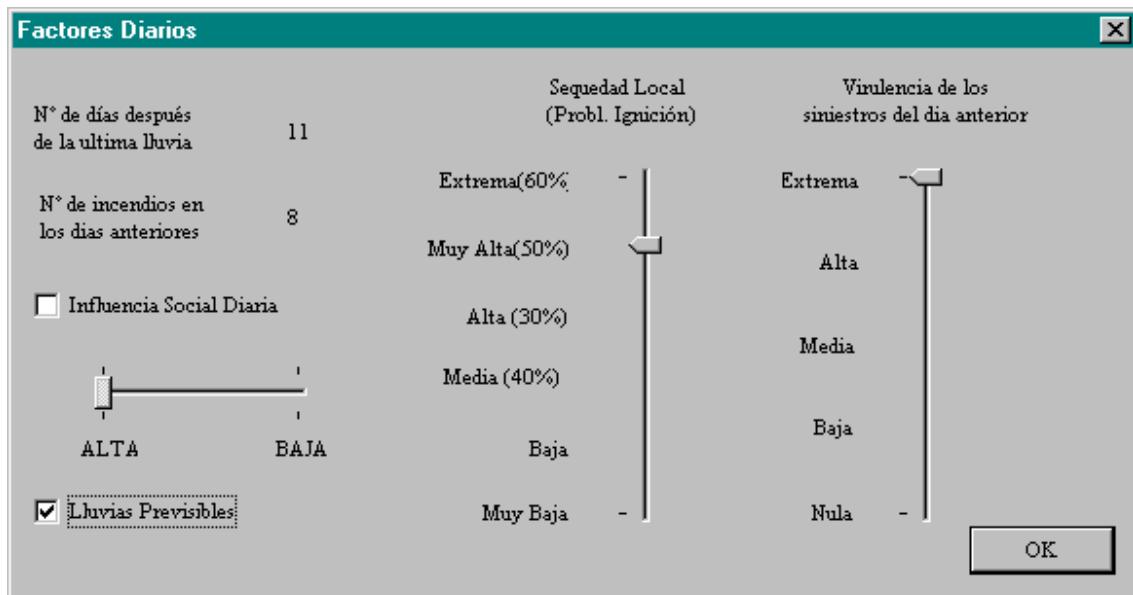


Figura 71: Factores diarios del 17 de Agosto.

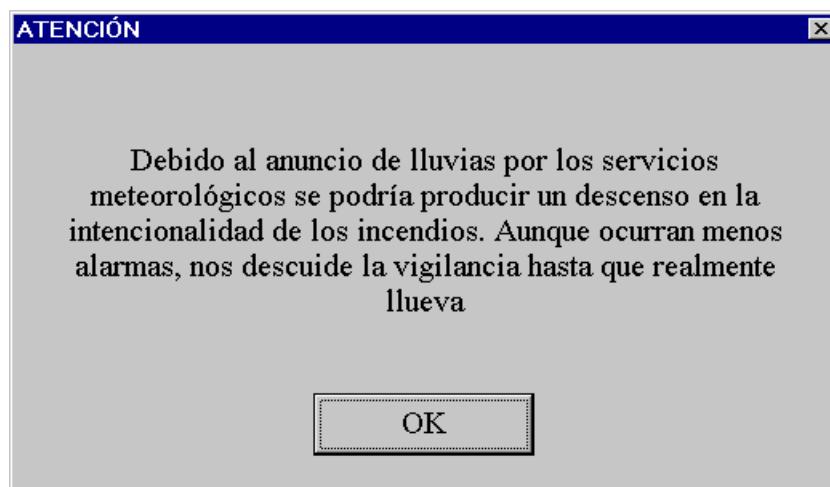


Figura 72: Mensaje de *Previsión de lluvias*.

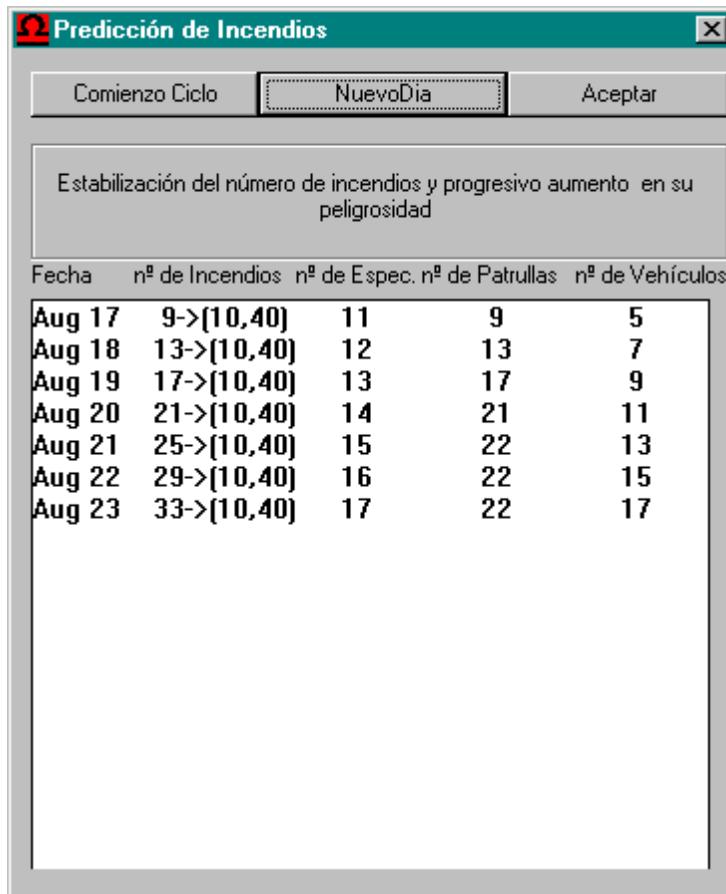


Figura 73: Predicción del 17 de Agosto.

Comparando lo ocurrido con lo que se predijo:

Día	NumIncendios	Especial	Personas	Vehículos
17-agosto-98	3	4	6	4
Predicción	9	11	9	5

Los datos que el usuario introdujo el día 18 fueron los siguientes. Nótese que sigue la previsión de lluvias dada por los servicios meteorológicos, por lo que el usuario recibe el mismo mensaje de aviso.

Factores Diarios			
Nº de días después de la ultima lluvia	12	Sequedad Local (Probl. Ignición)	Virulencia de los siniestros del dia anterior
Nº de incendios en los dias anteriores	3	Extrema(60%)  Muy Alta(50%)	Extrema  Alta Media Baja Nula
<input type="checkbox"/> Influencia Social Diaria		Alta (30%)  Media (40%)	
<input checked="" type="checkbox"/> Lluvias Previsibles		Muy Baja 	
		<input type="button" value="OK"/>	

Figura 74: Factores diarios del 18 de Agosto.

Comienzo Ciclo	NuevoDia	Aceptar		
Estabilización del número de incendios y progresivo aumento en su peligrosidad				
Fecha	nº de Incendios	nº de Espec.	nº de Patrullas	nº de Vehículos
Aug 18	5->[10,40]	10	5	3
Aug 19	9->[10,40]	11	9	5
Aug 20	13->[10,40]	12	13	7
Aug 21	17->[10,40]	13	17	9
Aug 22	21->[10,40]	14	21	11
Aug 23	25->[10,40]	15	22	13
Aug 24	29->[10,40]	16	22	15
Aug 25	33->[10,40]	17	22	17

Figura 75: Predicción del 18 de Agosto.

Comparando lo ocurrido con lo que se predijo:

Día	NumIncendios	Especial	Personas	Vehículos
18-agosto-98	4	8	4	4
Predicción	5	10	5	3

A última hora del día 18 comenzó a llover en toda la provincia de Lugo y se dio por finalizado el ciclo.

3.1.1. Análisis de los resultados de validación.

Se ha diseñado un programa que convierte en ficheros HTML los resultados obtenidos del estudio del comportamiento del modelo de predicción ante ciclos históricos representativos. Para el ciclo que se ha presentado, esta tabla sería:

Factores de ciclo														
Sequedad estacional				Media										
Intensidad Ultima Lluvia				Tormenta										
Influencia Social				Mucha										
Fecha	Seq.Local	Virulencia	Otros	Inc	Esp	Patr	Veh	Inc	Esp	Patr	Veh			
Aug 6	---	---	---	1	3	5	2	1	1	1	0			
Aug 07	Baja	Baja	---	3	4	6	3	1	1	2	1			
Aug 08	Baja	Baja	---	3	4	6	3	5	3	8	6			
Aug 09	Media	Media	---	6	6	5	4	4	6	8	4			
Aug 10	Media	Media	---	6	6	5	4	8	9	14	11			
Aug 11	Alta	Media	IFSB	16	10	13	10	16	27	25	39			
Aug 12	Alta	Alta	---	21	14	21	11	7	16	18	12			
Aug 13	Muy Alta	Alta	---	17	13	17	9	6	12	14	8			
Aug 14	Muy Alta	Alta	---	9	11	9	5	7	11	10	7			
Aug 15	Muy Alta	Alta	---	9	11	9	5	8	8	8	6			
Aug 16	Muy Alta	Extrema	IFSA	9	11	9	5	9	18	26	23			
Aug 17	Muy Alta	Extrema	LLU	9	11	9	5	3	4	6	4			
Aug 18	Alta	Alta	LLU	5	10	5	3	4	8	4	4			

Tabla 75: Resumen de los resultados reales y de predicción para el ciclo en

estudio.

Haciendo un análisis comparativo entre los valores reales y los que se han predecido para cada día, el resultado mostrado en los siguientes gráficos da muestra de que la tendencia de la predicción es la adecuada y que se producen disparidades cuando hay picos en la situación real, cuestión difícil de controlar.

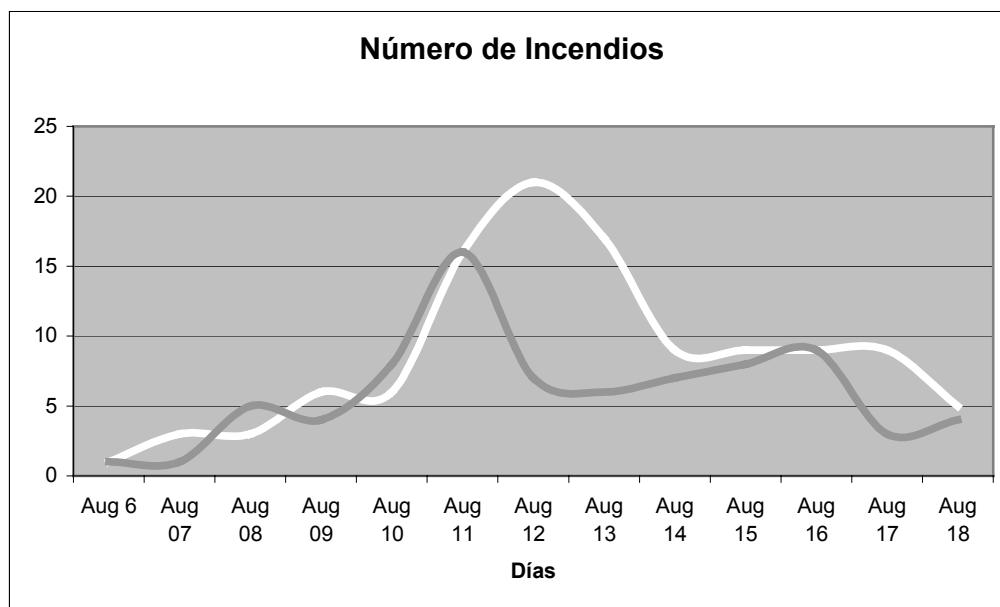


Figura 76: Resultados reales y de predicción para el ciclo en estudio en cuanto a número de incendios (predicción en blanco).

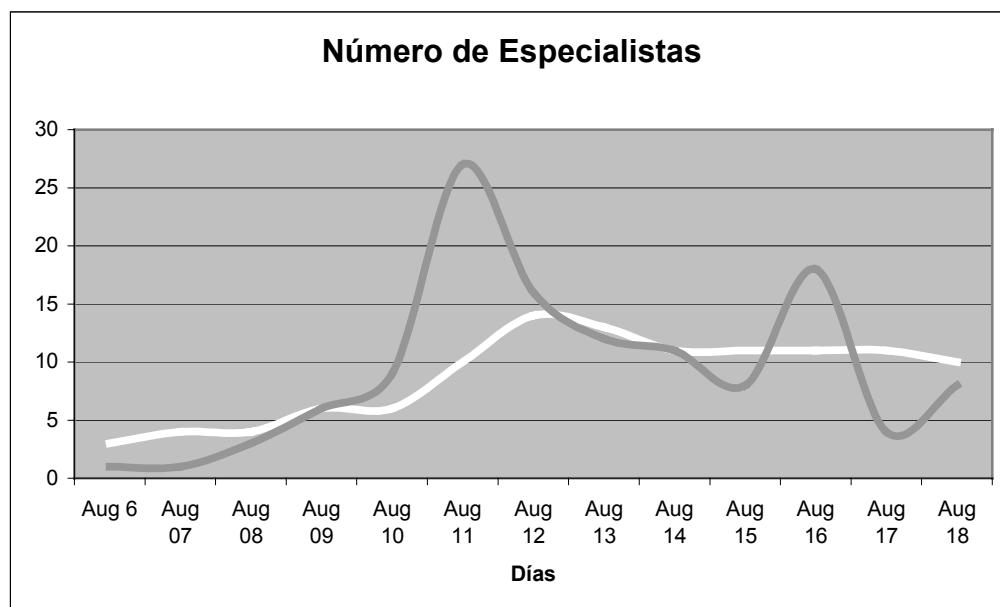


Figura 77: Resultados reales y de predicción para el ciclo en estudio en cuanto a número de especialistas (predicción en blanco).

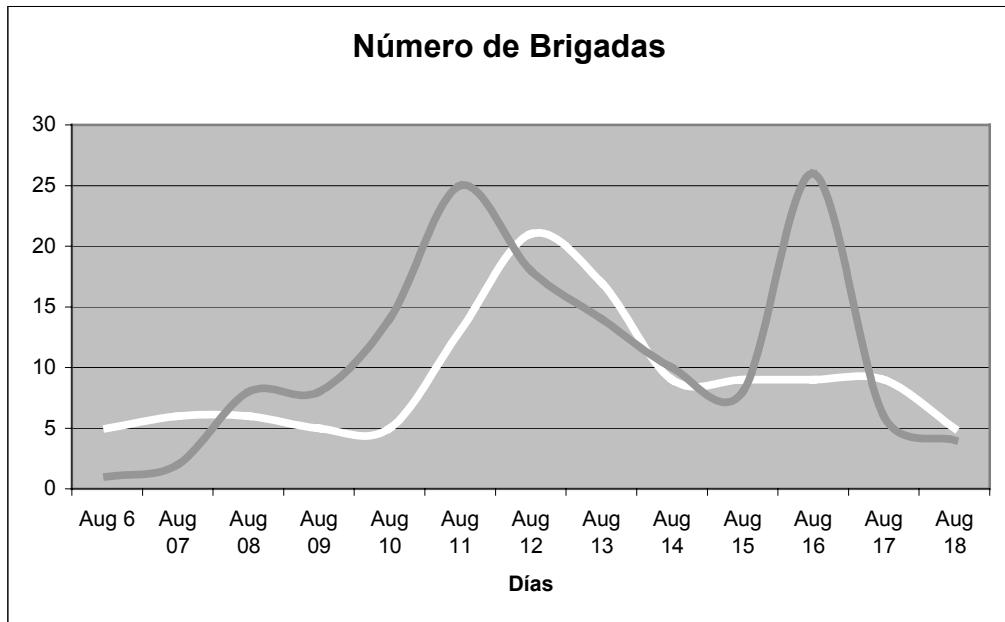


Figura 78: Resultados reales y de predicción para el ciclo en estudio en cuanto a número de brigadas (predicción en blanco).

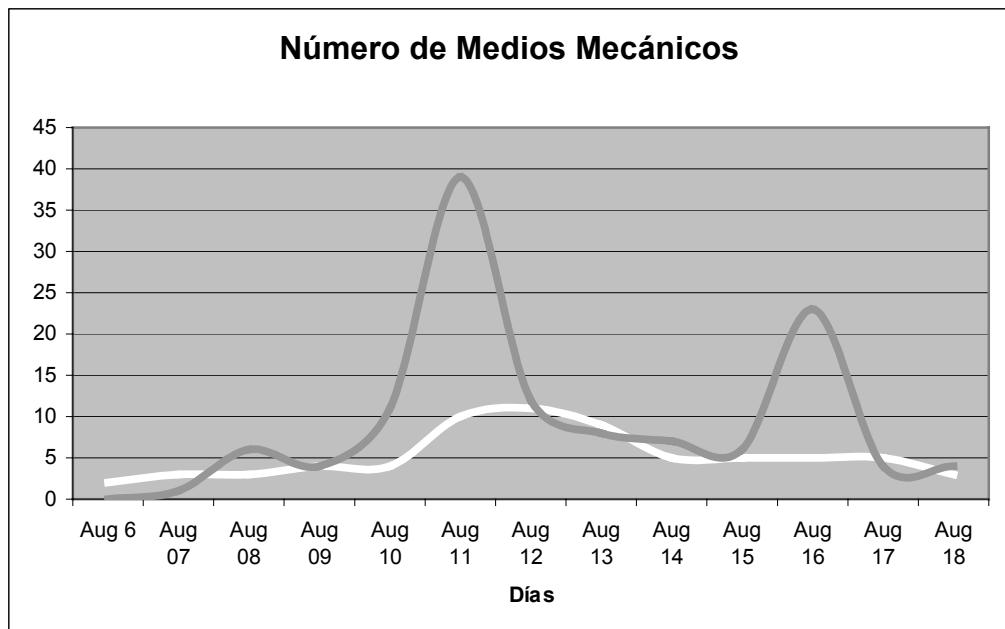


Figura 79: Resultados reales y de predicción para el ciclo en estudio en cuanto a número de medios mecánicos (predicción en blanco).

Si en vez de comparar los resultados reales con los de predicción puntuales se comparan con los intervalos de análisis (ver cap. 4), la respuesta del sistema es

mucho más exacta, pero estos intervalos no se muestran al usuario porque para él no serían útiles.

Por lo tanto, se puede concluir que los resultados de predicción del sistema son muy buenos, con una tasa de aciertos situada en torno al 80%. Pero lo mas importante es que estos resultados son útiles para ayudar a los usuarios del sistema en la toma de decisiones en su tarea diaria.

Capítulo 6

C onclusiones.

Se ha presentado un estudio experimental de la predicción basada en categorías deformables borrosas, que se ha implantado en INCEND-IA como Sistema Basado en el Conocimiento para la predicción y Ayuda a la Decisión en los Incendios Forestales.

La necesidad experimental de estudios de este tipo se justifica en la necesidad de contrastar los modelos teóricos con problemas reales complejos, ya que, habitualmente se utilizan casos de prueba de laboratorio que no tienen en cuenta el factor de escala. El desarrollo de interfaces, algoritmos, etcétera, para casos reales posibilita el acercamiento del Ingeniero de Conocimiento al problema, lo que conlleva expectativas y soluciones más realistas.

El tipo de problemas a los que se puede aplicar este método son aquellos de predicción que puedan ser descritos mediante patrones cíclicos o evolutivos en el tiempo. También es posible aplicarlo a situaciones estáticas, por ejemplo patrones de enfermedades con los que se compara una situación real (ejemplos en Apéndice 3), y se deforman los patrones afines para dar un diagnóstico y proporcionar un tratamiento.

Pero como conclusión parece adecuada una pequeña reflexión sobre determinadas cuestiones cruciales, tanto para las aportaciones de este trabajo como para el uso real al que está destinado el sistema.

1. ¿Qué problemas teóricos se han encontrado?. ¿Cómo se han resuelto?

1. Encontrar una noción de prototipo que permita manipular información cíclica-evolutiva, no esté representada por elementos concretos y las clases admitan solapamientos.

Se ha solucionado introduciendo una nueva noción basada en la idea de prototipo de Zadeh y en las investigaciones en el campo de la Psicología Cognitiva: Las Categorías Prototípicas Borrosas, que sí cumplen estos objetivos.

2. Descubrir estas Categorías Prototípicas Borrosas en una base de datos donde no están explícitas y en problemas complejos.

Se ha presentado una extensión del KDD (*Knowledge Discovery in Databases*), denominada Descubrimiento de Conocimiento Prototípico Borroso.

3. Formalizar estas Categorías Prototípicas Borrosas para poder ser utilizadas en un sistema informático real.

Se ha resuelto representando cada Categoría Prototípica Borrosa en una tabla-marco y asignándole un número borroso con el que se puede calcular la compatibilidad.

4. Deformar o desplazar estos prototipos para que modelizan una situación real.

Es decir, realicen el proceso inverso a lo que habitualmente se suele hacer: Asimilar el comportamiento de una situación real al de los prototipos.

Se ha resuelto introduciendo un nuevo concepto: El de Categorías Prototípicas Deformables Borrosas, basado en la noción de Categoría Prototípica Borrosa ya definida y dotándola de capacidad de deformación usando ideas del ámbito del reconocimiento de patrones (prototipos deformables, Bremermann).

5. Cómo comparar una situación real con las Categorías Prototípicas deformables Borrosas y cómo deformar éstas para que describan la situación real.

Se ha solucionado diseñando un proceso de equiparación de marcos basado en los grados de compatibilidad de la situación real con las Categorías Prototípicas Deformables Borrosas, utilizando, entre otras, técnicas de razonamiento no-monótono, en particular de razonamiento borroso no-monótono.

2. ¿Han encontrado útiles los usuarios los resultados del sistema?

Si, en las pruebas en paralelo realizadas se ha visto que puede ser muy útil el disponer de información como por ejemplo el número de brigadas que previsiblemente se necesitarán en los próximos días o como situar adecuadamente los recursos de los que se dispone para que una zona quede cubierta óptimamente.

Los usuarios han encontrado cómodo y comprensible el formato en el que se les presentan los resultados y no han tenido grandes dificultades para familiarizarse con los interfaces y el manejo del sistema.

Los técnicos, aunque habitualmente no son los usuarios directos del sistema (un operador es el que manipula directamente el ordenador), consideran valiosas las tablas de predicción y los consejos de redistribución de recursos y suelen utilizarlos en formato impreso.

3. ¿Provoca el sistema mejoras en la calidad de las decisiones?

Un sistema como INCEND-IA aplicado como ayuda de decisión en los centros de coordinación de lucha contra los incendios forestales de Galicia puede ser de gran ayuda debido a que:

- Colabora en la rapidez de las decisiones.
- Facilita una distribución de recursos más ordenada.
- Ayuda a evitar el efecto del nerviosismo y la precipitación ante situaciones conflictivas.
- Ayuda al personal del centro de coordinación a tomar decisiones ante una eventual ausencia de técnicos.

- Ayuda a la organización de recursos en días sucesivos.
- Una vez construido es fácilmente adaptable a todas las regiones.

4. ¿Qué aproximaciones han sido inefectivas en este problema?

Básicamente se han abordado tres aproximaciones alternativas para atacar el problema o alguna de sus partes.

- **Redes neuronales.** Se han probado redes autoorganizadas (mapas de Kohonen) para descubrir los prototipos, pero la búsqueda de Categorías Prototípicas Borrosas (muy complejas) ha hecho ineficiente este método. Para la predicción se han probado Perceptrones Multicapa con *backpropagation*, pero la naturaleza de la solución que debían reportar (ciclos dinámicos) los ha hecho ineficientes.
- **Algoritmos Genéticos.** El problema principal que se ha tenido al intentar usar computación evolutiva ha sido cómo codificar los datos de entrada en un problema tan complejo, por lo que se ha desecharido la idea.

5. ¿Qué métodos han sido usados para evaluar la efectividad del sistema?

Los resultados de INCEND-IA provocan una situación paradójica: Cuanto mejor es la predicción, peor es la similitud de los datos estimados con los reales. Esto se debe a que si se toman las medidas que aconseja el sistema, la ocurrencia de incendios disminuye en los siguientes días porque hay una mayor efectividad.

El sistema ha sido verificado y validado siguiendo las pautas propuestas en la metodología IDEAL, como se ha mostrado en el capítulo 5 de este trabajo.

Lo más significativo ha sido la validación, que consiste en la aseveración de que el sistema ofrece los resultados esperados y se comporta adecuadamente. Para ello se han hecho una serie de comprobaciones, se enuncian las más significativas y se ha obtenido un documento de validación:

Validación sobre el documento conceptual.

Se ha comprobado que los resultados del sistema son coherentes con las especificaciones del documento de conceptualización. Como ha existido alguna anomalía se ha corregido y realizado de nuevo la comprobación.

Validación sobre los casos históricos.

Ha consistido en comprobar los resultados obtenidos tomando como ejemplo los datos anteriores existentes, y haciendo las correcciones necesarias para su óptimo funcionamiento. Esta validación se hizo paralelamente a la anterior.

Validación sobre funcionamiento paralelo con casos reales.

Es lo que se podría llamar una fase de pruebas del sistema. Ha consistido en el uso del sistema en los casos reales en una oficina central de Lugo, estudiando las anomalías o fallos que se pudiesen dar y subsanándolos. Se realizó durante los meses de Julio, Agosto y Septiembre, que son los meses de campaña, normalmente con mayor número de incendios y más medios para combatirlos.

Bibliografía.

- [AAVV, 93a] AAVV Informe de la Comisión sobre Incendios Forestales. Senado. Madrid, 1993.
- [AAVV, 93b] AAVV Manual de operaciones contra incendios forestales. ICONA. Madrid, 1993.
- [AAVV, 95a] AAVV Informe de la Comisión sobre Incendios Forestales. Senado. Madrid, 1995.
- [AAVV, 95b] AAVV MAPA (1995).
- [AAVV, 96] AAVV Los incendios forestales en España durante el decenio 1986-95, MMA. Madrid, 1996.
- [AAVV, 97a] AAVV Los incendios forestales en España durante 1996. MMA. Madrid, 1997.
- [AAVV, 97b] AAVV Anuario de Estadística Agraria. MAPA. Madrid, 1997.
- [AAVV, 98a] AAVV Los incendios forestales en España durante 1997. MMA. Madrid, 1998.
- [AAVV, 98b] AAVV Directriz Básica de Planificación de Protección Civil de Emergencia por Incendios Forestales. Dirección General de Protección Civil, 1998.
- [Adán, 91] Adán, M. C.; Sobrino, A.: Nota sobre la teoría de prototipos. *Actas del Primer Congreso Español sobre Tecnologías y Lógica Fuzzy*. Universidad de Granada, 1991, pp. 173 - 177.
- [Adriaans, 95] Adriaans, P. W.: Predicting Pilot bid behavior with Genetic Algorithms. In *Proceedings of the Sixth International Conference on Human-Computer Interactions*. Tokyo, 1995.

- [Adriaans, 96] Adriaans, P. W.; Zantinge, D.: *Data Mining*. Addison-Wesley, 1996.
- [Agrawal, 96] Agrawal, R.; Mannila, H.; Srikant, R.; Toivonen, H.; Verkamo, A.: Fast Discovery of Association Rules. In *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press, Cambridge MA, 1996.
- [Aguado, 98] Aguado, I.; Chuvieco, E.; Camarasa, A.; Martín, P.; Camia, A.: Estimation of meteorological fire danger indices from multitemporal series of NOAA-AVHRR data. *III International Conference on Forest Fire Research. 14th Conference on Fire and Forest Meteorology*. Coimbra, Portugal, 1998, vol 1, pp. 1131 - 1147.
- [Aho, 88] Aho, A. V.; Hopcroft, J. E.; Ullman, J. D: *Estructuras de datos y algoritmos*. Addison-Wesley Iberoamericana. Wilmington, Delaware, E. U. A., 1988.
- [Alonso, 96] Alonso, M.; Camaras, A.; Chuvieco, E.; Kyun, I. A.; Martín, M. P.; Salas, F. J.: Estimating temporal dynamics of fuel moisture content of Mediterranean species from NOAA-AVHRR data, *EARSel Advances in Remote Sensing*, 1996, pp. 9 - 21.
- [Altrock, 94] Altrock, C. V.; Krause, B.: Multi-criteria decision making in German automotive industry using fuzzy logic. *Fuzzy Sets and Systems* 63, 1994, pp. 375 - 380.
- [Anderson, 91] Anderson, K. R.; Lee, B. L.: Linear programming techniques for initial attack resources deployment. *11th Conf. Fire and Forest Meteorology*, Missoula, Montana, 1991, pp. 1 - 8.
- [Anderson, 98] Anderson, K. R.; Hirsch, K.: A Preparedness Planning Evaluation Model. *III International Confer. on Forest Fire Research. 14th Conference on Fire and Forest Meteorology*. Coimbra, Portugal, 1998, vol. 2, pp. 2269 - 2278.
- [Arenas, 95] Arenas, L.: *Lógica formal para informáticos*. Díaz de Santos. Madrid, 1995.
- [Armstrong, 83] Armstrong, S. L.; Gleitman, L. R.; Gleitman, H.: What some Concepts might not be. *Cognition* 13, pp. 363 – 308, 1983.
- [Backer, 81] Backer, E.; Jain, A. K.: A Clustering Performance Measure Based on Fuzzy Set Decomposition. *IEEE Trans. Pattern Anal. Machine Intell.* PAMI-3 (1), 1981, pp. 66 - 75.
- [Barro, 98] Barro, S.; Bugarín, A.; Sobrino, A. (eds.): *Advances in Fuzzy Logic*. Universidad de Santiago de Compostela, 1998.
- [Batur, 91] Batur, C.; Kasparian, V.: Predictive fuzzy expert controllers. *Computer industrial engineering*, 20 (2), 1991, pp. 199 - 209.
- [Bellman, 70] Bellman, R.; Zadeh, L. A.: *Decision Making in a Fuzzy Environment*. Managemente Science, 17, 1970, pp. 141 - 174.
- [Bellman, 66] Bellman, R.; Kalaba, R.; Zadeh, L. A.: Abstraction and Pattern Classification. *Journal Math. Anal. Appl.* 13, 1966, pp. 1 - 7.
- [Berry, 96] Berry, M. J. A.; Linoff, G.: *Data Mining Techniques*. Wiley Computer Publishing. New York, 1996.

- [Besnard, 88] Besnard, P.; Siegel, P.: Preferential Models Approach to Non-Monotonic Logics. En Smets et al. (eds.) *Non-Standard Logics for Automated Reasoning*, Academic Press, 1988, pp. 137 - 156.
- [Bezdek, 77] Bezdek, J. C.; Castelaz, P. F.: Prototype Classification and Feature Selection with Fuzzy Sets. *IEEE Trans Syst., Man and Cybern.* SMC-7 (2), 1977, pp. 87 - 92.
- [Bezdek, 78] Bezdek, J. C.; Harris, J. D.: Fuzzy Partitions and relations: An Axiomatic Basis for Clustering. *Fuzzy Sets and Systems* 1, 1978, pp. 111 - 127.
- [Bezdek, 80] Bezdek, J. C.: A Convergence Theorem for the Fuzzy ISODATA Clustering Algorithms. *IEEE Trans. Pattern Anal. Machine Intell.* PAMI-2 (1), 1980, pp. 1 - 8.
- [Bezdek, 85] Bezdek, J. C.; Anderson, I. M.: An Application of the c-Varieties Clustering Algorithms to Polygonal Curve Fitting. *IEEE Trans Syst., Man and Cybern.* SMC-15 (5), 1985, pp. 637 - 641.
- [Bezdek, 87] Bezdek, J. C.; Hathaway, R. J.; Sabin, M. J.; Tucker, W. T.: Convergence Theory for Fuzzy c-Means Counterexamples and Repairs. *IEEE Trans Syst., Man and Cybern.* SMC-17 (5), 1987, pp. 873 - 877.
- [Bezdek, 91] Bezdek, J. C.; Pal, S. K. (eds.): *Fuzzy Models for Pattern Recognition (Methods that search for structures in data)*. IEEE Press, 1991.
- [Bonissone, 82] Bonissone, P. P.: A Fuzzy Sets Based Linguistic Approach: Theory and Applications. En M. M. Gupta y E. Sanchez (eds.) *Approximate Reasoning in Decision Analysis*, North-Holland Publishing Company, 1982, pp. 329 - 339.
- [Bonissone, 84] Bonissone, P. P.; Tong R. M.: *Linguistic solutions to fuzzy decision problems*. TIMS/Studies in Management Sciences, 20. Elsevier Science Pub, 1984, pp. 323 - 334.
- [Bonissone, 86] Bonissone, P. P.; Decker, K. S.: Selecting Uncertainty Calculi and Granularity: An Experiment in Trading-off Precision and Complexity. En L. N. Kanal y J. F. Lemmer (eds.) *Uncertainty in Artificial Intelligence*, Elsevier Science Publishers, 1986, pp. 217 - 247.
- [Boolos, 89] Boolos, G. S.; Jeffrey, R. S.: *Computability and logic*. Cambridge University Press. Cambridge, 1989.
- [Bossu, 85] Bossu, G.; Siegel, P.: Saturation, non Monotonic Reasoning and The Closed World Assumption. *Artificial Intelligence* 25 (1), 1985, pp. 13 - 65.
- [Brassard, 90] Brassard, G.; Bratley, P.: *Algorítmica: Concepción y análisis*. Masson. Barcelona, 1990.
- [Bratko, 86] Bratko, I.: *Prolog programming for artificial intelligence*. Addison-Wesley. Reading, Mass., 1986.
- [Breiman, 84] Breiman, L.; Friedman, J. H.; Olshen, R. A.; Stone, C. J.: *Classification and Regression Trees*. Wadsworth, Belmont, CA, 1984.
- [Bremermann, 76] Bremermann, H.: Pattern Recognition. H. Bossel: *Systems Theory in the Social Sciences*. Birkhäuser Verlag, 1976, pp. 116 - 159.

- [Bruner, 56] Bruner, J. S.; Goodnow, J. S.; Austin, G. A.: *A Study of Thinking*. Wiley, 1956.
- [Buchanan, 84] Buchanan, B. G.; Shortliffe, E. H. (eds.): *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Addison-Wesley, Reading, Mass. 1984.
- [Caballero, 98] Caballero, D.: FOMFIS: A Computer Based System for Forest Fire Prevention Planning. *III International Confer. on Forest Fire Research. 14th Conference on Fire and Forest Meteorology*. Coimbra, Portugal, 1998, vol 2, pp. 2643 - 2652.
- [Caetano, 94] Caetano, M. S.; Mertes, L. A. K.; Pereira, J. M. C.: Using Spectral Mixture analysis for fire severity mapping. *2th Int. Conf. Forest Fire Research*. Coimbra, Portugal, 1994, vol 2(16), pp. 667 - 677.
- [Carbonell, 90] Carbonell, J.: *Machine Learning, Paradigms and Methods*. MIT Press, Cambridge MA, 1990.
- [Casanova, 98] Casanova, J. L.; Calle, A.; González-Alonso: A forest Fire Risk Assessment obtained in Real time by means of NOAA satellite images. *III International Confer. on Forest Fire Research. 14th Conference on Fire and Forest Meteorology*. Coimbra, Portugal, 1998, vol 1, pp. 1169 - 1171.
- [Castro, 93] Castro, J.L.; Trillas, E.: The management of the inconsistency in expert systems. *Fuzzy Sets and Systems* 58, North Holland, 1993, pp. 51 – 57.
- [Castro, 94] Castro, J.L.; Trillas, E.: Constraints as incompatibility relations in KBS. *International Journal of Uncertainty, Fuzziness and KBS*, Vol. 2, No. 1, 1994, pp. 121 – 126.
- [Castro, 98] Castro, J.L.; Trillas, E.; Zurita, J. M.: Non-monotonic Fuzzy Reasoning. *Fuzzy Sets and Systems* 94, North Holland, 1998, pp. 217 - 225.
- [Chang, 77] Chang, R. L. ; Paulidis, T.: Fuzzy Decision Tree Algorithms. *IEEE Trans Syst., Man and Cybern. SMC-7* (1), 1977, pp. 28 - 35.
- [Chen, 98] Chen, J.; Xi, Y.; Zhang, Z.: A clustering algorithm for fuzzy model identification. *Fuzzy Sets and Systems* 98 (3), 1998, pp. 319 - 329.
- [Chepoia, 99] Chepoia, V.; Dumitrescu, D.: Fuzzy clustering with structural constraints. *Fuzzy Sets and Systems* 105 (1), 1999, pp. 91 - 97.
- [Clark, 78] Clark, K.: Negation as Failure. En H. Gallaire y J. Minker (eds.) *Logic and Data Bases*, Plenum Press, 1978, pp. 293 - 322.
- [Clark, 96] Clark, P.; Niblett, T.: The CN2 Induction Algorithm. *Machine Learning*, 3, 1996, pp. 261 – 283.
- [Cohen, 77] Cohen, G.: *The Psychology of Cognition*. Academic Press, 1977.
- [Collado, 87] Collado, M.; Morales, R.; Moreno, J. J.: *Estructuras de Datos*. Díaz de Santos. Madrid, 1987.
- [Cooper, 61] Cooper, C. F.: La ecología del fuego. En *Ecología, evolución y biología de poblaciones*. Selecciones de Scientific American. Omega. Barcelona, 1961, pp. 294 - 302.

- [Cox, 98] Cox, E.: *The Fuzzy Systems Handbook* (Second Edition). Academic Press, 1998.
- [Cuena, 86] Cuena, J. (ed.): *Inteligencia Artificial: Sistemas Expertos*. Alianza Editorial. Madrid, 1986.
- [De Groot, 91] De Groot, W. J.: Development of Saskatchewan's Fire Suppression Preparedness System. Proceedings of the *6th Central Region Fire Weather Committee Scientific and Technical Seminar*. Winnipeg, Manitoba, 1991, No. 5, pp: 23 - 49.
- [De las Heras, 91] De las Heras, J.; Martínez, J. J.; Herranz, J. M.: Impacto Ecológico de los Incendios Forestales. *Revista de Estudios Albacetenses* 29, 1991, pp. 105 - 117.
- [Deaño, 94] Deaño, A.: *Introducción a la lógica formal*. Alianza Universidad Textos. Madrid, 1994.
- [Delgado, 96a] Delgado, M.; Gómez Skarmeta, A. F.; Jiménez, L.: Regression by fuzzy knowledge bases. *Proc. EUFIT'96* 2, 1996, pp. 1170 - 1176.
- [Delgado, 96b] Delgado, M.; Gómez Skarmeta, A. F.; Jiménez, L.: Arboles difusos de clasificación y regresión. *Actas VI Congreso Español sobre Tecnologías y Lógica Fuzzy*, 1996, pp. 101 - 106.
- [Delgado, 98] Delgado, M.; Gómez Skarmeta, A. F.; Martín, F.: A methodology to model fuzzy systems using fuzzy clustering in a rapid-prototyping approach. *Fuzzy Sets and Systems* 97 (3), 1998, pp. 287 - 301.
- [Deshayes, 98] Deshayes, M.; Chuvieco, E.; Cocero, D.; Karteris, M.; Koutsias, N.; Stach, N.: Evaluation of different NOAA-AVHRR derived indices for fuel moisture content estimation interest for short term fire risk assessment. *III International Confer. on Forest Fire Research. 14th Conference on Fire and Forest Meteorology*. Coimbra, Portugal, 1998, vol 1, pp. 1149 - 1167.
- [Devi, 86] Devi, B. B.; Sarma, V. V. S.: A Fuzzy Approximation Scheme for Sequential Learning in Pattern Recognition. *IEEE Trans Syst., Man and Cybern. SMC-16* (5), 1986, pp. 668 - 679.
- [Dimitrescu, 88] Dimitrescu, D.: Hierarchical Pattern Classification. *Fuzzy Sets and Systems*, 28, 1988, pp. 145 - 162.
- [Di Natale, 95] Di Natale C.; Davide F.; D'Amico A.: Complex Chemical pattern recognition with sensor array : the discrimination of vintage years of wine. Department of Electronic Engineering University of Rome "Tor Vergata". *Sensors and Actuators B*, 24-25, 1995, pp 801 - 804.
- [Dimitrakopoulos, 98] Dimitrakopoulos, A.P.; Yordanov, G.: A Computer Program for Forest Fire Data Inventory and Statistical Analysis. *III International Confer. on Forest Fire Research. 14th Conference on Fire and Forest Meteorology*. Coimbra, Portugal, 1998, vol 2, pp. 2491 - 2502.
- [Dubois, 80] Dubois, D.; Prade, H.: *Fuzzy Sets and Systems: Theory and Applications*. Academic Press, New York, 1980.

- [Duda, 79] Duda, R.; Gaschnig, J.; Hart, P.: Model design in the Prospector consultant system for mineral exploration. En Michie D. (ed.): *Expert Systems in the Microelectronic Age*. Edinburgh University Press, 1979, pp. 153 - 167.
- [Duncker, 45] Duncker, K.: On problem solving. *Psychol. Monogr.*, vol 58(5), nº 270, 1945.
- [Dunn, 73] Dunn, J. C.: A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal Cybernetics* 3 (3), 1973, pp. 32 - 57.
- [Dunn, 74] Dunn, J. C.: A Graph Theoretic Analysis of Pattern Classification via Tamura's Fuzzy Relation. *IEEE Trans Syst., Man and Cybern.* SMC-4 (3), 1974, pp. 310 - 313.
- [Dyckhoff, 84] Dyckhoff, H.; Pedrycz, W.: Generalized Means as a Model of Compensative Connectives. *Fuzzy Sets and Systems* 14, 1984, pp. 143 - 154.
- [Eftichidis, 98] Eftichidis, G.; Varela, V.; Margaritis, E.: Prometheus System: A modern approach for Wildfire Management in the Mediterranean Ecosystems. *III International Conference on Forest Fire Research. 14th Conference on Fire and Forest Meteorology*. Coimbra, Portugal, 1998, vol 2, pp. 2349 - 2350.
- [Eidenshink, 90] Eidenshink, J. C.; Burgan, R. E.; Haas R. H.: Monitoring fire fuels condition by using time series composites of Advanced Very High Resolution Radiometer (AVHRR) Data. Proc. Resource Technology 90. ASPRS, Washington, D.C., 1990, pp. 68 - 82.
- [FAO, 92] FAO. Global wildland fire statistics, 1981-1990. MISC/92/4. Forestry Department, Roma, 1992.
- [Fayyad, 96a] Fayyad, U.; Piatetsky-Shapiro, G.; Smyth, P.: The KDD Process for Extracting Useful Knowledge from Volumes of Data. *Communications of the ACM*, November 1996/ Vol 39, Nº 11, pp. 27 – 34.
- [Fayyad, 96b] Fayyad, U.; Piatetsky-Shapiro, G.; Smyth, P.; Uthurusamy, R. (Eds): *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press, Cambridge MA, 1996.
- [Fernández, 87] Fernández, G.; Sáez Vacas, F.: *Fundamentos de Informática*. Alianza-Informática. Madrid, 1987.
- [Friedman, 97] Friedman, M.; Henneb, M.; Kandela, A.: Most typical values for fuzzy sets. *Fuzzy Sets and Systems* 87 (1), 1997, pp. 27 - 37.
- [Fu, 98] Fu, G.: Optimization methods for fuzzy clustering. *Fuzzy Sets and Systems* 93 (3), 1998, pp. 301 - 309.
- [Fuhrmann, 91] Fuhrmann, GY.: Note on the Integration of Prototype Theory and Fuzzy-set Theory. *Synthese* 86, 1991, pp. 1 - 27.
- [García, 97] García, C.: Programas Comunitarios de I+D en Incendios Forestales. *Estado de la Investigación y el desarrollo en la protección contra incendios forestales en España, I Seminario Nacional, Ponencias*. Lugo, 1997, pp. 1 - 12.

- [Gardner, 92] Gardner J. W.; Hines E. L.; Tang H. C.: Detection of Vapours and odours from a multisensor array using pattern-recognition techniques. Department of Engineering, University of Warwick, Coventry. *Sensors and Actuators B*, 9, 1992, pp. 9 - 15.
- [Garrido, 95] Garrido, M.: *Lógica simbólica*. Tecnos. Madrid, 1995.
- [Gath, 89] Gath, I.; Geva, A. B.: Unsupervised Optimal Fuzzy Clustering. *IEEE Trans. Pattern Anal. Machine Intell.* PAMI-11 (7), 1989, pp. 773 - 781.
- [Genesereth, 87] Genesereth, M. R.; Nilsson, N. J.: *Logical foundations of artificial intelligence*. Morgan Kaufmann. Los Altos CA, 1987.
- [Genther, 97] Genther, H.; Glesner, M.: Advanced data preprocessing using fuzzy clustering techniques. *Fuzzy Sets and Systems* 85 (2), 1997, pp. 155 - 164.
- [Gitman, 70] Gitman, I.; Levine, M. D.: An Algorithm for Detecting Unimodal Fuzzy Sets and Its Application as a Clustering Technique. *IEEE Trans. Comput.* C-19 (7), 1970, pp. 583 - 593.
- [Glymour, 96] Glymour, C; Madigan, D.; Pregibon, D.; Smyth, P.: Statistical Inference and Data Mining. *Communications of the ACM*, November 1996/ Vol 39, Nº 11, pp. 35 – 41.
- [Godó, 89] Godó, L.; López de Mántaras, R.; Sierra, C.; Verdaguer, A.: MILORD: the architecture and management of linguistically expressed uncertainty. *International Journal of Intelligent Systems* 4, 1989, pp. 471 - 501.
- [Gómez, 97] Gómez, A.; Juristo, N.; Montes C.; Pazos, J.: *Ingeniería del Conocimiento*. Centro de Estudios Ramón Areces, Madrid, 1997.
- [Gómez-Skarmeta, 99] Gómez Skarmeta, A. F.; Delgado, M.; Vila, M. A.: About the use of fuzzy clustering techniques for fuzzy model identification. *Fuzzy Sets and Systems* 106 (2), 1999, pp. 179 - 188.
- [González-Alonso, 97a] González-Alonso, F.; Cuevas, J. M.; Casanova, J. L.; Calle, A.; Illera, P.: A forest fire risk assessment using NOAA AVHRR images in the Valencia area, eastern Spain. *Int. J. Remote Sensing* 18(10), 1997, pp. 2201 - 2207.
- [González-Alonso, 97b] González-Alonso, F.; Ustin, S. L.; Palacios-Orueta, A.; Gardner, M.: Utilización de imágenes AVIRIS en la Cartografía del combustible en el Chaparral de California. *Ecología* 11, 1997, pp. 31 - 36.
- [González-Alonso, 98] González-Alonso, F.: Aplicaciones de la Teledetección en el estudio de los incendios forestales. *Mapping*, 1998, pp. 14 - 20.
- [Gustafson, 79] Gustafson, D. E.; Kessel, W. C.: Fuzzy clustering with a fuzzy covariance matrix. *Proc. IEEE CDC*, San Diego CA, 1979, pp. 761 - 766.
- [Hirota, 96] Hirota, K.; Pedrycz, W.: Directional fuzzy clustering and its application to fuzzy modelling. *Fuzzy Sets and Systems* 80 (3), 1996, pp. 315 - 326.
- [Hirota, 97] Hirota, K.; Pedrycz, W.: Nonmonotonic Fuzzy Set Operations: A Generalization and Some Applications. *International Journal of Intelligent Systems*, vol. 12, 1997, pp. 483 - 493.

- [Hirsch, 91] Hirsch, K. G.: Development of an initial attack preparedness system for Manitoba. *11th Conf. Fire and Forest Meteorology*, Missoula, Montana, 1991, pp:16 - 19.
- [Hunt, 66] Hunt, E. B.; Marin, J.; Stone, P. J.: *Experiments in Induction*, Academic Press, New York, 1966.
- [Hunt, 75] Hunt, E. B.: *Artificial Intelligence*, Academic Press, New York, 1975.
- [Huntsberger, 90] Huntsberger, T. L.; Ajjimarangsee, P.: Parallel self-organizing feature maps for unsupervised pattern recognition. *International Journal of General Systems* 16 (4), 1990, pp. 357 - 372.
- [Iancu, 98] Iancu, I.: Propagation of Uncertainty and Imprecision in Knowledge-Based Systems. *Fuzzy Sets and Systems* 94, North Holland, 1998, pp. 29 - 43.
- [Jiménez, 97] Jiménez, L.: *Modelización difusa de sistemas mediante técnicas inductivas*. Tesis Doctoral, Universidad de Granada, Julio de 1997.
- [Jones, 82] Jones, G. V.: Stacks not Fuzzy Sets: An Ordinal Basis of Prototype Theory of Concepts. *Cognition* 12, 1982, pp. 281 - 290.
- [Keller, 85a] Keller, J. M.; Gray, M. R.; Givens, J. A.: A Fuzzy K-Nearest Neighbor Algorithm. *IEEE Trans Syst., Man and Cybern. SMC-15* (4), 1985, pp. 580 - 585.
- [Keller, 85b] Keller, J. M.; Hunt, D. J.: Incorporating Membership Functions into the Perceptron Algorithm. *IEEE Trans. Pattern Anal. Machine Intell. PAMI-7* (6), 1985, pp. 693 - 699.
- [Klir, 88] Klir, G. J.; Folger, T. A.: *Fuzzy Sets, Uncertainty and Information*. Prentice Hall, 1988.
- [Kodratoff, 88] Kodratoff, Y.: *Introduction to Machine Learning*. Pitman, London, 1988.
- [Kohonen, 84] Kohonen, T.: *Self-Organization and Associative Memory*. Springer-Verlag, Berlín, 1984.
- [Kowalski, 79] Kowalski, R.: *Logic for problem solving*. North Holland. Amsterdam, 1979.
- [Lago, 93] Lago, R.; Olivas, J. A.; Sobrino, A.: Distribución de medios en INCEND-IA mediante técnicas de Lógica Borrosa. S.Barro and A. Sobrino (eds.): *III Congreso Español sobre Tecnologías y Lógica Fuzzy. Comunicaciones*. Universidad de Santiago de Compostela, 1993, pp. 107 - 113.
- [Lago, 92] Lago, R.; Olivas, J. A.; Sobrino, A.: Una posible aplicación de la teoría de prototipos de Zadeh a la gestión de la evolución de la siniestralidad de los incendios forestales en un motor de inferencia. *Actas del II Congreso Español sobre Tecnologías y Lógica Fuzzy*. Facultad de Informática, U.P.M, 1992, pp. 47 - 57.
- [Lago, 96] R. Lago R.; Olivas J. A.; J. A. Megido J. A.; Grávalos C.; Pérez J.: INFEDEC 2.1: Sistema de ayuda en la decisión de diagnóstico y tratamiento de Enfermedades Infecciosas. Actas de *ESTYLF'96: VI Congreso Español sobre Tecnologías y Lógica Fuzzy*, Universidad de Oviedo, 1996, pp. 271 - 275.

- [Lasaponara, 98] Lasaponara, R.; Cuomo, V.; Tramutoli, V.: Satellite forest fire detection in the italian ecosystems usin AVHRR data. *III International Confer. on Forest Fire Research. 14th Conference on Fire and Forest Meteorology.* Coimbra, Portugal, 1998, vol 2, pp. 2013 - 2028.
- [Lavrac, 94] Lavrac, N.; Dzeroski, S.: *Inductive Logic Programming: Techniques and Applications*. Ellis Horwood, Chichester UK, 1994.
- [Lavrac, 95] Lavrac, N.; Wrobel S. K. (Eds): *Machine Learning: ECML-95*. Springer-Verlag, New York, 1995.
- [Lee, 75] Lee, S. C.; Lee, E. T.: Fuzzy Neural Networks. *Mathematical Biosciences* 23, 1975, pp. 151 - 177.
- [Lifschitz, 85] Lifschitz, V.: Computing Circumscription. *Proc. 9th International Joint Conference on Artificial Intelligence (IJCAI'85)*, Morgan Kaufmann, Los Angeles, 1985.
- [Lifschitz, 86] Lifschitz, V.: Pointwise Circumscription. *Proc. AAAI*, 1986.
- [Lim, 90] Lim, Y. W.; Lee, S. U.: On the color image segmentation algorithm based on the theresholding an the fuzzy c-means techniques. *Pattern Recognition* 23 (9), 1990.
- [Linares, 99] Linares, L. J.; Olivas, J. A.: Representación de Conceptos Difusos mediante Indices Semánticos". *Proceedings of the EUSFLAT - ESTYLF Joint Conference 1999*. September 22-25, Palma de Mallorca, Spain, pp. 441 - 444.
- [Lindsay, 80] Lindsay, R. K.; Buchanan, B. G.; Feigembaum, E. A.; Lederberg, J.: *Applications of Artificial Intelligence for Organic Chemistry: The DENDRAL Project*. McGraw-Hill, New York, 1980.
- [López, 91] López, S.; González, F.; Llop, R.; Cuevas, M.: An evaluation of the utility of NOAA-AVHRR images for monitorin forest fire risk in Spain. *International Journal of Remote Sensing* 12, 1991, pp. 1841 - 1851.
- [López de Mántaras, 90] López de Mántaras, R.: *Approximate Reasoning Models*. Ellis Horwood series in artificial Intelligence, Chichester UK, 1990.
- [Luchins, 42] Luchins, A. S.: Mechanization in the problem solving: The effect of Einstellung. *Psychology Monogr.*, vol. 54(6), nº 248, 1942.
- [Martell, 89a] Martell, D. L.; Otukol, S.; Stocks, B. J.: A logistic model for predicting daily people-caused forest fire occurrence in Ontario. *Canadian Journal Forestal Research* 17, 1989, pp. 394 - 401.
- [Martell, 89b] Martell, D. L.; Bevilacqua, E.: Modelling seasonal variation in daily people-caused forest fire occurrence. *Canadian Journal Forestal Research* 19, 1989, pp. 1555 - 1563.
- [Martín, 94] Martín, M. P.; Viedma, O.; Chuvieco, E.: High versus low resolution satellite images to estimate burned areas in large forest fires. *2th Int. Conf. Forest Fire Research*. Coimbra, Portugal, vol. 2(14), 1994, pp. 653 - 663.
- [Martín-Rico, 97] Martín-Rico, C.: Detección y monitorización operacional desde satélite. Programa Fuego, *Estado de la Investigación y el desarrollo en la protección contra incendios forestales en España, I Seminario Nacional, Ponencias*. Lugo, 1997, pp. 269 - 274.

- [Martínez, 98] Martínez, E.: *Sistemas de detección*. Curso Básico de Especialización en Defensa Contra Incendios Forestales. COITF, Albacete, 1998.
- [Maté, 88] Maté, J. L., Pazos, J.: *Ingeniería del conocimiento*. SEPA. Córdoba, Argentina, 1988.
- [Mc Carthy, 86] Mc Carthy, J.: Applications of Circumscription to Formalizing Commonsense Knowledge. *Artificial Intelligence* 28, 1986, pp. 89 - 116.
- [Mérida, 98] Mérida, J. C.: *Nuevas tecnologías en la defensa contra incendios forestales*. Curso Básico de Especialización en Defensa Contra Incendios Forestales. COITF, Albacete, 1998.
- [Mervis, 81] Mervis, C. B.; Rosch, E.: Categorization of Natural Objects. *Ann. Rev. Psychology* 32, 1981, pp. 89 -115.
- [Michalski, 94] Michalski, R.; Tecuci, G.: *Machine Learning: A Multistrategy Approach, Volume IV*. Morgan Kaufmann, Redwood City CA, 1994.
- [Michalewicz, 94] Michalewicz, Z.: *Genetic Algorithms + Data Structures = Evolution Programs*. Springer-Verlag, New York, 1994.
- [Michie, 96] Michie, D.; Speigelhalter, D. J.; Taylor C.C.: *Machine Learning, Neural and Statistical Classification*. Ellis Horwood, Chichester UK, 1996.
- [Minsky, 75] Minsky, M.: A Framework for Representing Knowledge. En P. Winston (ed.) *The Psychology of Computer Vision*, McGraw-Hill, 1975, pp. 211 - 277.
- [Montero de Burgos, 87] Montero de Burgos, J. L.: La regresión vegetal y la restauración forestal. *Bol. Est. Cent. Ecol.* 31, 1987, pp. 5 - 22.
- [Montoya, 86] Montoya, J. M.: Ecología y pinares. *Bol. Est. Cent. Ecol.* 30, 1986, pp. 19 - 24.
- [Moreno, 98] Moreno, J. M.; Vazquez, A.; Velez, R.: Recent history of forest fires in Spain. En: *Large forest fires*. Backhuys Publishers, Leiden, Holanda, 1998, pp. 159 - 185.
- [Morik, 89] Morik, K. Ed.: *Knowledge Representation and Organisation in Machine Learning*. Springer-Verlag, New York, 1989.
- [Mukaidono, 87] Mukaidono, M.; Shen, Z.; Ding, L.: *Fuzzy Prolog*. Proc. 2nd. IFSA Congress, Tokyo, Japan, 1987, pp. 452 -455.
- [Nath, 83] Nath, A. K.; Lee, T. T.: On the design of a classifier with linguistic variables as inputs. *Fuzzy Sets and Systems* 11, 1983, pp. 265 - 286.
- [Negoita, 85] Negoita, C. V.: *Expert Systems and Fuzzy Systems*. Benjamin/Cummings, Menlo Park, CA, 1985.
- [Olivas, 91] Olivas, J. A.; Sobrino A.: Aproximación Lingüística y teoría de prototipos de Zadeh. Actas del *Primer Congreso Español sobre Tecnologías y Lógica Fuzzy*, Universidad de Granada, 1991, pp. 179 - 183
- [Olivas, 92] Olivas, J. A.: El tratamiento del lenguaje impreciso en los Sistemas Expertos. Inchaurrealde C., Olivares C. & Navarro P. (Eds.): *Semántica y Lenguajes especializados*. Universidad de Zaragoza, 1992, pp. 127 - 137.

- [Olivas, 93] Olivas, J. A.; Sobrino A.; Lago, R.: INCEND-IA: Un prototipo de sistema experto para la previsión y asignación de medios en los incendios forestales. Bustos, E. de, Echeverría, J., Pérez Sedeño, E. y Sánchez Balmaseda, M.I. (eds.): *Actas del I Congreso de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España*. Departamento de Reprografía de la Uned, 1993, pp. 84 - 87.
- [Olivas, 94] Olivas, J. A.; Sobrino, A.: An Application of Zadeh's Prototype Theory to the Prediction of Forest Fire in a Knowledge-based System. *Proceedings of the 5th. International IPMU* (Information Processing and Management of Uncertainty in Knowledge-based Systems), VOL. II, Paris, 1994, pp. 747 - 752.
- [Olivas, 95a] Olivas, J. A.; Sobrino A.: An application of fuzzy sets theory to the prediction of forest fire in a knowledge-based system. Proceedings of the *International Conference Forest Fires: Initiation, spread and ecological impacts*, Tomsk State University, Tomsk, Siberia, Russia, 1995, pp. 76 - 77.
- [Olivas, 95b] Olivas, J. A.; Sobrino A.: Una aplicación del concepto de prototipo deformable a un método que gestiona incertidumbre en un SBC. *Anales de Informática VOLUMEN 1, ESTYLF'95: Tecnologías y Lógica Fuzzy*, Facultad de Informática, Universidad de Murcia, 1995, pp. 303 - 308.
- [Olivas, 97a] Olivas, J. A.; Sobrino A.; Criado, J. M.; de Agapito, J. A.: Una aplicación de la Teoría de Prototipos de Zadeh a la Gestión del Reconocimiento de la aparición de Incendios Forestales utilizando Sensores de Estado Sólido. *CIMAF'97: Simposio de Inteligencia Artificial*. La Habana, Cuba, 1997, pp. 209 - 210.
- [Olivas, 97b] Olivas, J. A.: Método de Predicción de incendios mediante Lógica Fuzzy. *Estado de la Investigación y el desarrollo en la protección contra incendios forestales en España, I Seminario Nacional, Ponencias*. Lugo, 1997, pp. 69 - 80.
- [Olivas, 98] Olivas, J. A.; Ordax, J. M.; Ocaña, A.: Lógica Borrosa en Internet: Implementación en JAVA de Sistemas Expertos de Formación. *3º Simpósio de Investigação e Desenvolvimento de Software Educativo*, Evora, Portugal, 1998. <http://www.minerva.uevora.pt>. Actas, p. 19.
- [Olivas, 99a] Olivas, J. A.; Sobrino, A.; Prieto, M.; Vizcaíno, A.; Lanza, S. F.: Fuzzy Prototypes for Prediction and Diagnosis: Two Applications. *Proceedings of ISAS'99: Second International Symposium on Artificial Intelligence* (ICIMAF, Institute of Cybernetics, Mathematics and Physics, La Habana, Eds: A. Ochoa, M. R. Soto y R. Santana), 1999, pp. 134 - 142.
- [Olivas, 99b] Olivas, J. A.: Applications of Fuzzy Prototypes: Forest Fire Prediction and Medical Diagnosis. *Proceedings of the International Conference on Artificial Intelligence IC-AI'99*. CSREA Press, USA (Computer Science Research, Education and Applications Press), Ed. H. R. Arabnia, 1999, pp. 662 - 668.
- [Olivas, 00] Olivas, J. A.: Aplicación de Lógica Borrosa y Prototipos a la predicción de incendios forestales, diagnóstico de enfermedades infecciosas y reconocimiento de olores. En *Líneas de Investigación en Informática*, Coordinadores: J. Bravo y M. Ortega. *Ciencia y Técnica 26*. Ediciones de la Universidad de Castilla-La Mancha, 2000. pp. 143 - 153.

- [Olivas, 00] Olivas, J. A.; Romero, F. P.: "FPKD: Fuzzy Prototypical Knowledge Discovery. Application to Forest Fire Prediction". *Proceedings of the SEKE'2000*, Knowledge Systems Institute, Chicago, Ill. USA, 2000, pp. 47 - 54.
- [Ollero, 98] Ollero, A.; Martínez de Dios, J. R.; Arrúe, B. C.: Integrated Systems for Early Forest-Fire Detection. *III International Confer. on Forest Fire Research. 14th Conference on Fire and Forest Meteorology*. Coimbra, Portugal, 1998, vol. 2, pp. 1977 - 1988.
- [Osheron, 81] Osheron, D. N.; Smith, E. E.: On the adequacy of Prototype Theory of Concepts. *Cognition* 9, 1981, pp. 35 – 58.
- [Osheron, 82] Osheron, D. N.; Smith, E. E.: Gradedness and Conceptual Combination. *Cognition* 12, 1982, pp. 299 – 318.
- [Pedernera, 96] Pedernera, P.: Modelo de optimización para el diseño de Sistemas de Torres de Detección de Incendios Forestales. *Chile Forestal* (100). CONAF, 1996.
- [Pathak, 86] Pathak, A; Pal, S. K.: Fuzzy Grammars in Syntactic Recognition of Skeletal Maturity from X-Rays. *IEEE Trans Syst., Man and Cybern. SMC-16* (5), 1986, pp. 657 - 667.
- [Pedrycz, 98] Pedrycz, W.: Fuzzy set technology in knowledge discovery. *Fuzzy Sets and Systems* 98 (3), 1998, pp. 279 - 290.
- [Peneva, 98] Peneva, V.; Popchev, I.: Comparison of clusters from fuzzy numbers. *Fuzzy Sets and Systems* 97 (1), 1998, pp. 75 - 81.
- [Peña, 93] Peña, R.: *Diseño de programas: Formalismo y abstracción*. Prentice Hall International (UK), 1993.
- [Pérez, 99] Pérez, E.: Minería de Datos: Fundamentos, Métodos y Aplicaciones. *Novatica, nº 138*, 1999, pp. 36 - 44.
- [Peterson, 79] Peterson, P.: On the Logic of Few, Many and Most. *Notre Dame Journal of Formal Logic*, vol. 20, 1979, pp. 159 - 179.
- [Piatetsky, 91] Piatetsky-Shapiro, G.; Frawley, W.: *Knowledge Discovery in Databases*. AAAI/MIT Press, Cambridge MA, 1991.
- [Pinilla, 95] Pinilla, C.: *Elementos de Teledetección*. RA-MA. Madrid, 1995.
- [Popper, 59] Popper, K.: *The Logic of Scientific Discovery*. Hutchinson, London, 1959.
- [Priber, 97] Priber, U.; Kretzschmar, W.: Inspection and supervision by means of hierarchical fuzzy classifiers. *Fuzzy Sets and Systems* 85 (2), 1997, pp. 263 - 274.
- [Quinlan, 79] Quinlan, J. R.: Discovering Rules by Induction from a Large Collection of Examples. En D. Michie (ed.) *Expert Systems in the Microelectronic Age*, Edinburgh University Press, 1979.
- [Quinlan, 83] Quinlan, J. R.: Induction of Decision Trees. En *Machine Learning* 1, 1983, pp. 81 - 106.
- [Quinlan, 88] Quinlan, J. R.: *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo CA, 1988.

- [Quiñoá, 97] Quiñoá, J. L.; Porto, R.: Asignación de recursos en la lucha contra incendios forestales. En: *I Seminario Nacional Sobre Estado de la Investigación y el Desarrollo en protección contra Incendios Forestales en España*. Lugo, 1997, pp. 297 - 306.
- [Rallo, 97] Rallo, M.: Alta Tecnología de Defensa en Defensa del Medio Ambiente. El Proyecto Bosque. En: *I Seminario Nacional Sobre Estado de la Investigación y el Desarrollo en protección contra Incendios Forestales en España*. Lugo, 1997, pp. 205 - 211.
- [Reiter, 80] Reiter, R.: A Logic for Default Reasoning. *Artificial Intelligence* 13, 1980, pp. 81 -132.
- [Rips, 95] Rips, L. J.: The Current Status of Research on Concept Combination. *Mind & Language* 10, N° 1/2 March/June 1995, pp. 72 - 104.
- [Roberts, 93] Roberts, D.; Adams, J. B.; Smith, M. O.: Discriminating green vegetation, non photosynthetic vegetation and soils in AVIRIS data. *Remote Sensing of Environment* 44, 1993, pp. 1 - 25.
- [Rosch, 75] Rosch, E.: Cognitive Representations of Semantic Categories. *Journal of Experimental Psychology, General* 104, pp. 192 – 233, 1975.
- [Rosch, 76] Rosch, E.; Mervis, C. B.; Gray, W. D.; Johnson, D. M.; Boyes-Braem, P.: Basic objects in natural categories, en *Cognit. Psychol.* 8, 1976, pp. 382 - 439.
- [Rosch, 77] Rosch, E.: Clasification of real-world objects: origins and representations in cognition, en P. N. Johnson-Laird y P. C. Wason (eds.), *Thinking: Reading in Cognitive Science*, Cambridge University Press, 1977.
- [Roth, 83] Roth, E. M.; Mervis, C. B.: Fuzzy Set Theory and Class Inclusion Relations in Semantic Categories. *Journal of Verbal Learning and Verbal Behavior* 22, 1983, pp. 509 - 525.
- [Roubens, 78] Roubens, M.: Pattern classification problems and fuzzy sets. *Fuzzy Sets and Systems* 1, 1978, pp. 239 - 253.
- [Ruiz, 97] Ruiz, J. R.; Castaño, S.; Calera, A.: Aplicaciones de la Teledetección y los SIG a la determinación del riesgo de erosión en zonas incendiadas. *I Congreso Forestal Hispano-Luso*. Irati. Navarra, 1997, vol. 5, pp. 425 - 430.
- [Rumelhart, 86] Rumelhart, D.; McClelland, J.: *Parallel Distributed Programming, Volumes I and II*. MIT Press, Cambridge MA, 1986.
- [Ruspini, 69] Ruspini, E.: A New Approach to Clustering. *Inform. and Control* 15, 1969, pp. 22 - 32.
- [Sabin, 87] Sabin, M. J.: Convergence and Consistency of Fuzzy c-Means/ISODATA Algorithms. *IEEE Trans. Pattern Anal. Machine Intell.* PAMI-9 (5), 1987, pp. 661 - 668.
- [Sahnoun, 91] Sahnoun, Z.; DiCesare, F.; Bonissone, P. P.: Efficient methods for computing linguistic consistency. *Fuzzy Sets and Systems* 39, North-Holland, 1991, pp. 15 - 26.
- [Schank, 77] Schank, R. C.; Abelson, R. P.: *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures*. Erbaum, Hillsday, NJ. 1977.

- [Shoham, 87] Shoham, Y.: A Semantical Approach to Non Monotonic Logics. En Ginsberg, M. L. (ed.) *Readings in Non Monotonic Reasoning*, 1987.
- [Silgeström, 95] Silgeström, P.; Moreno, A.: Monitoring burnt areas by principal components analysis of multi-temporal TM data. *Int. J. Remote Sensing*. 16(9), 1995, pp.1577 - 1587.
- [Silva, 98] Silva, F. R.: Mathematical model to determine the dimensions of the protection infrastructures in the techniques of preventive silviculture. *III International Confer. on Forest Fire Research. 14th Conference on Fire and Forest Meteorology*. Coimbra, Portugal, 1998, vol. 2, pp. 2255 - 2268.
- [Simon, 95] Simon H. A.: Artificial Intelligence: An empirical science. *Artificial Intelligence*, 77, 1995, pp. 95 - 127.
- [Siy, 74] Siy, P.; Chen, C. S.: Fuzzy logic for Handwritten Numerical Character Recognition. *IEEE Trans Syst., Man and Cybern.*, 1974, pp. 570 - 575.
- [Smith, 81] Smith, E. E.; Medin, D. L.: *The Psychology of Conceptual Processes*. Harvard Univ. Press. Cambridge, Mass. 1981.
- [Smith, 83] Smith, E. E.; Osheron, D. N.: Conceptual Combination with Prototype Concepts. *Cognitive Science* 8, pp. 337 – 361, 1983.
- [Sobrino, 92] Sobrino, A.: Is there a prototypical logic for prototypes and classification founded on them?. *Proceedings of the IPMU'92*, Mallorca, 1992, pp. 697 - 704.
- [Sobrino, 93a] Sobrino A.; Olivas, J. A.; Adán, M. C.: Temas de Lógica Borrosa y Razonamiento Aproximado: Pasado, Presente y Proyecciones Futuras (Capítulo I). S. Barro y A. Sobrino (eds.): *Estudios de lógica borrosa y sus aplicaciones*. Universidad de Santiago de Compostela, 1993, pp. 13 - 48.
- [Sobrino, 93b] Sobrino A.; Olivas, J. A.; Adán, M. C.: Comentarios sobre la evolución de las máquinas Fuzzy-Prolog. S. Barro y A. Sobrino (eds.): *III Congreso Español sobre Tecnologías y Lógica Fuzzy*. Comunicaciones. Universidad de Santiago de Compostela, 1993, pp. 97 - 106.
- [Sobrino, 93c] Sobrino A.; Barro, S. (eds.): *Estudios de lógica borrosa y sus aplicaciones*. Universidad de Santiago de Compostela, 1993.
- [Sobrino, 94] Sobrino A.; Olivas, J. A.: Semioraciones y semicadenas (Estudio de algunos aspectos). Francesc Esteva y Pere García (eds.): *Tecnologías y Lógica Fuzzy*. Institut d'Investigació en Intelligència Artificial - CSIC, 1994, pp. 139 - 145.
- [Sobrino, 95] Sobrino, A.; Olivas, J. A.: Gramáticas de estructuras de frase y expresividad de los conceptos vagos. *Anales de Informática VOLUMEN I, ESTYLF'95: Tecnologías y Lógica Fuzzy*, Facultad de Informática, Universidad de Murcia, 1995, pp. 129 - 134.
- [Sobrino, 96] Sobrino, A.; Olivas, J. A.: Phrase Structure Grammars for the Expression of Vague Concepts in Spanish. *Mathware & Soft Computing* 3, 1996, pp.395 - 402.
- [Sugeno, 85a] Sugeno, M.; Nishida, M.: Fuzzy Control of Model Car, *Fuzzy Sets and Systems* 16, North-Holland, 1985, pp. 103 - 113.

- [Sugeno, 85b] Sugeno, M. (ed.): *Industrial Applications of Fuzzy Control*. North-Holland, 1985.
- [Tai, 98] Tai, W. C.; Goldgof, D. G.; Hall, L. O.: Fast fuzzy clustering. *Fuzzy Sets and Systems* 93 (1), 1998, pp. 49 - 56.
- [Takagi, 85] Takagi, T.; Sugeno, M.: Fuzzy Identification of Systems and its Applications to Modeling and Control. *IEEE Trans. On Systems, Man and Cybernetics* 15 (1), 1985.
- [Takagi, 91] Takagi, T.; Hayashi, I.: NN-Driven Fuzzy Reasoning. *International Journal of Approximate Reasoning* 5 (3), 1991, pp. 191 - 212.
- [Tamura, 71] Tamura, S.; Higuchi, S.; Tanaka, K.: Pattern Classification Based on Fuzzy Relations. *IEEE Trans Syst., Man and Cybern. SMC-1* (1), 1971, pp. 61 - 66.
- [Tong, 84] Tong, M.; Bonissone, P.: *Linguistic Solutions to Fuzzy Decision Problems*. TIMS/Studies in the Management Sciences, 20, 1984, pp. 323 - 334.
- [Torra, 97] Torra, V.: The Weighted OWA Operator. *International Journal of Intelligent Systems*, 12, 1997, pp. 153 - 166.
- [Trillas, 92] Trillas, E.; Gutierrez, J. (eds.): *Aplicaciones de la Lógica Borrosa*. Eds CSIC, Nuevas Tendencias 20, Madrid, 1994.
- [Trillas, 94] Trillas, E. (ed.): *Fundamentos e introducción a la ingeniería Fuzzy*. Omron Electronics S. A., Madrid, 1992.
- [Trillas, 95] Trillas, E.; Alsina, C.; Terricabras, J. M.: *Introducción a la Lógica Borrosa*. Ariel, Barcelona, 1995.
- [Tversky, 77] Tversky, A.; Kahneman, D.: Judgment under uncertainty: heuristics and bases, en P. N. Johnson-Laird y P. C. Wason (eds.), *Thinking: Reading in Cognitive Science*, Cambridge University Press, 1977.
- [Ullman, 89] Ullman, J. D.: *Principles of Database and Knowledge-base Systems*. Computer Science Press CA, 1989.
- [Vélez, 91] Vélez, R.: El problema de los incendios forestales. En: *Curso superior sobre defensa contra incendios forestales*. E.U.P. Albacete, 1991.
- [Vélez, 95] Vélez, R.: El peligro de incendios forestales derivado de la sequía. *Deterioro de los montes y cambio climático*. Sociedad Española de Ciencias Forestales. Madrid, 1995, pp. 99 - 109.
- [Vélez, 96] Vélez, R.: Instrumentos tecnológicos para el control de los incendios forestales. En: *Curso sobre Detección y Ayuda a la Decisión Aplicadas a Incendios Forestales*. INSA, 1996, pp. 3 - 12.
- [Vélez, 98] Vélez, R.: El período 1848-1997 en la defensa contra los incendios forestales en España. *Montes* 53, 1998, pp. 51 - 62.
- [Vizcaíno, 98a] Vizcaíno, A.; Prieto, M.; Olivas, J. A.: Synchronous Distributed Knowledge Construction for Learning. *3º Simpósio de Investigação e Desenvolvimento de Software Educativo*, Evora, Portugal, 1998. <http://www.minerva.uevora.pt>. Actas, p. 14.

- [Vizcaíno, 98b] Vizcaíno, A.; Prieto, M.; Olivas, J. A.: APIC: A tool for Cooperative Knowledge Construction and Learning. *3º Simpósio de Investigação e Desenvolvimento de Software Educativo*, Evora, Portugal, 1998. <http://www.minerva.uevora.pt>. Actas, p. 13.
- [Vizcaíno, 99] Vizcaíno, A.; Prieto, M.; Olivas, J. A.: Modelos del Estudiante en Entornos Colaborativos. *Actas de TISE 98 Taller Internacional de Software Educativo*, Santiago, Chile, 2, 3, y 4 de Diciembre de 1998.
- [Vizcaíno, 99] Vizcaíno, A.; Prieto, M.; Olivas, J. A.: Modelos del estudiante en entornos de Aprendizaje Colaborativo. *Actas del Congreso Internacional de Informática Educativa 99*. UNED, 1999, pp. 19 - 26.
- [Weiss, 91] Weiss, S.; Kulikowski, C.: *Computer Systems that Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning and Expert Systems*. Morgan Kaufmann, Redwood City CA, 1991.
- [White, 96] White, J. D.; Ryan, K. C.; Key, C. C.; Running, S. W.: Remote Sensing of Forest Fire Severity and Vegetation Recovery. *Int. J. Wildland Fire* 6(3), 1996, pp. 125 - 136.
- [Windham, 82] Windham, M. P.: Cluster Validity for the Fuzzy c-Means Clustering Algorithm. *IEEE Trans. Pattern Anal. Machine Intell.* PAMI-4 (4), 1982, pp. 357 - 363.
- [Windham, 83] Windham, M. P.: Geometrical fuzzy clustering algorithms. *Fuzzy Sets and Systems* 10, 1983, pp. 271 - 279.
- [Winograd, 72] Winograd, T.: Understanding natural language. *Cognit. Psychol.* 3, 1972, pp. 1 - 191.
- [Winograd, 72] Winograd, T.: Frame representations and the declarative-procedural controversy. En *Representation and Understanding: Studies in Cognitive Science* (Bobrow y Collins, eds.), New York Academic, 1975, pp. 185 - 210.
- [Wisniewski, 96] Wisniewski, E. J.: Construal and Similarity in Conceptual Combination. *Journal of Memory and Language* 35, 1996, pp. 434 - 453.
- [Wisniewski, 98] Wisniewski, E. J.; Bradley, C. L.: Relations versus Properties in Conceptual Combination. *Journal of Memory and Language* 38, 1998, pp. 177 - 202.
- [Woodworth, 35] Woodworth, R. S.; Sells, S. B.: An atmosphere effect in formal syllogistic reasoning. *Journal of Experimental Psychology*, vol. 18, pp. 451 – 460, 1935.
- [Wybo, 98] Wybo, J. L.; Eftichidis, G.; Koutsouris, D.; Manganas, T.; Viegas, D. X.; Apostolopoulos, T.; Pelosio, E.; Bovio, G.; Ollero, A.; Schmidt, D.; Criado, A.: DEDICS: A General Framework for Supporting Management of Forest Fires. *III International Confer. on Forest Fire Research. 14th Conference on Fire and Forest Meteorology*. Coimbra, Portugal, 1998.
- [Xie, 91] Xie, X. L.; Beni, G.: A Validity Measure for Fuzzy Clustering. *IEEE Trans. Pattern Anal. Machine Intell.* PAMI-13 (8), 1991, pp. 841 - 847.
- [Yager, 87] Yager, R. R.: Using Approximate Reasoning to Represent Default Knowledge. *Artificial Intelligence* 31, 1987, pp. 99 - 112..

- [Yager, 88] Yager, R. R.: On Ordered Weighted Averaging Aggregation Operators in Multicriteria Decisionmaking. *IEEE Transactions on Systems, Man and Cybernetics*, vol. 18, nº 1, 1988, pp. 183 - 190.
- [Yager, 91] Yager, R. R.: Non-monotonic set theoretic operations. En *Fuzzy Sets and Systems* 42, North-Holland, 1991, pp. 173 - 190.
- [Yager, 97] Yager, R. R.: A Note on a Fuzzy Measure of Typicality. En *International Journal of Intelligent Systems* 12, John Wiley & Sons, 1997, pp. 233 - 249.
- [Yang, 96] Yang, M.; Koa, C.: On a class of fuzzy c-numbers clustering procedures for fuzzy data. *Fuzzy Sets and Systems* 84 (1), 1996, pp. 49 - 60.
- [Yang, 99] Yang, M.; Liua, H.: Fuzzy clustering procedures for conical fuzzy vector data. *Fuzzy Sets and Systems* 106 (2), 1999, pp. 189 - 200.
- [Zadeh, 73] Zadeh, L. A.: Outline of a new approach to the analysis of complex systems and decision processes, *IEEE Trans. On Systems, Man and Cybernetics* 3 (1), 1973, pp. 28 - 44.
- [Zadeh, 82] Zadeh, L. A.: A note on prototype set theory and fuzzy sets. *Cognition* 12, 1982, pp. 291-297.
- [Zadeh, 87] Zadeh, L. A.: *Fuzzy Sets and Applications* (Selected Papers, edited by R. R. Yager, S. Ovchinnikov, R. M. Tong, H. T. Nguyen), John Wiley, Nueva York, 1987.

Apéndice 1

M etodología, Conceptos y Herramientas.

1. Método de trabajo.

Por lo general, el desarrollo de un Sistema Basado en el Conocimiento suele ser, en parte, similar al de otros programas convencionales. Por ello, se podría decir que hay una parte donde surgen problemas y dificultades relacionadas con el contexto de los SBC y otra parte en la que las dificultades son comunes al desarrollo convencional de software. No obstante, se va a hacer hincapié en el ámbito que nos ocupa, y así tratar de describir las sucesivas fases que integran el proceso de desarrollo de un SBC, fases que al estar tan interrelacionadas entre sí, no siempre es posible establecer una clara separación entre las mismas. No existe una metodología única de desarrollo para los distintos tipos de Sistemas Basados en Conocimiento. Sin embargo, puede resultar interesante la elaboración de un esquema genérico aproximado, cuya estructura y orden podría variar dependiendo fundamentalmente del objetivo del programa a diseñar.

En el desarrollo de un SBC siempre se tiene en cuenta la interacción entre los Ingenieros de Conocimiento y los expertos humanos en el dominio. Ahora bien, la actuación de ambos no debería ignorar las aspiraciones y necesidades de una tercera entidad en juego que no es otra que los destinatarios o usuarios finales del programa. El papel protagonista del desarrollo suele corresponder a los dos primeros, pero el éxito o fracaso del sistema dependerá en gran medida de las aportaciones y colaboraciones que se hayan dado entre las tres entidades involucradas. continuación se describen brevemente las fases genéricas para el desarrollo de un SBC:

1.1. *Etapa 1: Definición del problema. Identificación.*

En un primer momento, previo al inicio del proceso de desarrollo de la aplicación informática, es sumamente importante realizar una descripción lo más detallada posible de la cuestión que se va a intentar resolver. Es muy conveniente indicar la misión, objetivos, cuáles van a ser las entradas al sistema, y especificar claramente las salidas del mismo. Es decir, se necesita aclarar desde un principio la situación de partida y el nivel de detalle al que se desea llegar.

Si esta fase no queda definida y acotada con suficiente claridad, es posible que al finalizar el desarrollo el usuario del sistema no quede satisfecho con el mismo, porque lo que está obteniendo como resultados, no le sirven para nada. Por tanto, si desde el principio, se tiene una visión muy clara de cuál es el problema y qué resultados se desean obtener, todo será mucho más fácil tanto para el usuario como para el equipo encargado de realizar el SBC.

1.2. *Etapa 2: Estudio de viabilidad.*

Lo que se pretende es comprobar si realmente el SBC va a poder contribuir, de una manera eficiente, a la resolución de un problema permitiendo alcanzar todos los objetivos previstos. No siempre la mejor alternativa para resolver convenientemente un problema tiene que ser de forma obligada un SBC, o expresándolo de una manera más genérica, las técnicas de Inteligencia Artificial. Son múltiples los métodos que se pueden aplicar con el fin de realizar este estudio, pero quizás uno de los más sencillos y precisos es el *método o test de Slagel* (1989) [Maté et al. 88].

1.3. *Etapa 3: Adquisición del conocimiento.*

Aunque no de forma exclusiva, casi siempre que dentro de este ámbito se hace referencia expresa a las posibles fuentes de las que se extraerá la experiencia e información, se suele pensar en un especialista o experto como protagonista fundamental del proceso. De hecho no parece factible plantear la creación de un programa que acumule experiencia y conocimientos pertenecientes a un campo específico del saber sin tener en cuenta, desde el principio, las fuentes de información provenientes del o los expertos. La elección de un buen experto como fuente inicial de conocimiento es de suma importancia ya que cuanto mayor sea la calidad de su experiencia mejor tenderá a ser la calidad de los resultados del futuro SBC.

No obstante, se puede disponer del mejor experto del mundo en una materia pero de nada sirve eso si carece de la capacidad y transparencia necesarias para la transmisión de ideas, argumentos, opiniones, creencias, razonamientos, pensamientos, intuiciones, etc. El Ingeniero del Conocimiento tendrá entonces que ingeníárselas para poder sonsacar la información que necesita del experto. Existen varias técnicas para conseguirlo.

1.4. *Etapa 4: Representación del conocimiento.*

Suele ser ésta una de las tareas que más tiempo y esfuerzo demandan de las personas que se dedican al diseño de SBCs. Ello se debe al hecho de no conocer aún con la suficiente precisión cuáles son los procesos que cualquier ser humano y, en particular un experto, activa en su mente cuando selecciona, examina, sintetiza y transforma los datos iniciales de un problema en pos de alcanzar una solución válida para el mismo.

La función de cualquier esquema de representación es capturar los rasgos esenciales del ámbito correspondiente a un problema concreto y hacer accesible esta información a un procedimiento de resolución específico. Como se sabe, diferentes tipos de problemas requieren diferentes tipos de razonamiento. A su vez, cada modalidad de razonamiento precisa de una adecuada representación del conocimiento.

1.5. *Etapa 5: Implementación.*

En esta fase se lleva a cabo el desarrollo íntegro del SBC en términos de programación. Se suelen construir prototipos que serán confrontados con el experto y rediseñados en numerosas ocasiones. Así se sigue una metodología incremental y cíclica con la que el sistema se va refinando hasta conseguir optimizarlo. También es importante decidir el lenguaje de programación utilizado para la implementación dependiendo de factores como los siguientes: Requerimientos de tiempos de respuesta, requerimientos de Interfaz de usuario, flexibilidad que se quiere aportar a la herramienta, requerimientos de hardware, coste en mantenimiento de la herramienta y facilidades aportadas al usuario.

1.6. ¿Qué procedimiento práctico se ha seguido para construir la aplicación INCEND-IA?

Siguiendo fundamentalmente la metodología IDEAL de construcción de Sistemas expertos, las fases de desarrollo de INCEND-IA se pueden esquematizar en las siguientes:

Adquisición de Conocimiento: En esta fase se ha tratado de adquirir el conocimiento necesario para abordar el problema, para posteriormente analizarlo e implementarlo en INCEND-IA.

Definición del problema: Antes de comenzar la adquisición del conocimiento propiamente dicha se ha planteado cuál va a ser el problema que se pretende resolver y hasta donde se quiere que llegue INCEND-IA en su resolución. Como resultado de esta fase se ha generado el Documento de alcance con sus anexos (no presente en esta memoria).

Adquisición del conocimiento experto: Consiste en la realización de una serie de entrevistas a los expertos con el objeto de obtener la información necesaria para resolver el problema planteado. Para ello se han usado técnicas propias de adquisición del conocimiento, como pueden ser los emparrillados o las entrevistas estructuradas. Cada entrevista se ha plasmado en un documento de adquisición de cada sesión.

Conceptualización y análisis: Una vez obtenido el conocimiento necesario, en esta fase se ha definido un diccionario de términos, unas tablas objeto-atributo-valor y algunas otras cosas por las que posteriormente se va a regir el sistema, que figuran en un documento de conceptualización. También se ha establecido un mapa de procesos y conocimientos, que describen los diferentes pasos y estados por los que se pasa para extraer una determinada conclusión.

Diseño del sistema formal: El objetivo de esta fase ha sido estructurar el conocimiento buscando un sistema adecuado de representación así como las posibles herramientas apropiadas para la implementación del prototipo, lenguajes de programación adecuados a lo que se tiene que implementar, como se va a organizar el control del sistema (metarreglas, pizarra, agenda...), así como definir su flujo de control, estrategias de búsqueda, heurísticas... También ha sido necesario dejar claro qué se necesita en los diferentes interfaces de comunicación. Ha sido una de las fases más críticas del proceso de construcción del sistema.

Entorno: El entorno de representación del conocimiento es clave para la futura implementación del sistema. En esta fase se ha necesitado pasar los datos obtenidos en la conceptualización a estructuras formales directamente implementables.

Herramientas: Esta fase se ha realizado al mismo tiempo que la anterior y consiste en la elección de una o varias herramientas para la construcción de Sistemas Expertos, de inducción, etc. así como los lenguajes de programación adecuados. No se han utilizado herramientas específicas para la construcción de Sistemas Expertos y se ha construido un sistema “a medida” utilizando el entorno Visual Estudio C++, así como todas las técnicas propias del entorno WINDOWS y herramientas necesarias por ejemplo para comunicar y gestionar bases de datos como ODBC y Access.

Interfaz: Es decisiva para el buen funcionamiento del sistema una comunicación agradable y eficaz con el usuario, por ello la presentación del prototipo y del sistema final se han diseñado utilizando el entorno WINDOWS, por su facilidad de manejo y lo práctico y agradable que resulta. También se han especificado claramente cuáles son los datos que el sistema necesita del exterior, así como de otras posibles fuentes (como puede ser una base de datos). Se ha establecido comunicación entre diferentes módulos del sistema.

Implementación del sistema: En esta fase se ha llevado a cabo la programación del sistema para su funcionamiento. Posteriormente se ha Verificado y Validado el sistema INCEND-IA.

Modificaciones del interfaz de usuario: Modificaciones en el interfaz de usuario, atendiendo a las propias experiencias y quejas de los operadores, para lograr que fuese lo más amistoso posible.

Elaboración de la documentación del sistema INCEND-IA: La documentación que acompaña al sistema final debe ser completa, comprensible para los usuarios y posibles futuros administradores del sistema. Está compuesta por los siguientes documentos: Documento de alcance y anexos, documentos de adquisición del conocimiento, documento de conceptualización, mapas de procesos

y conocimientos, documento de implementación, documento manual de uso del sistema y documento de interfaces. Toda esta documentación está integrada en un documento final del prototipo de INCEND-IA.

Gestión del proyecto: Se llevó a cabo durante toda la realización del proyecto y consistió en la administración de los recursos, el almacenamiento de documentos, recopilación de datos, planificación de entrevistas, etc.

2. Adquisición de conocimiento y *Data Mining*.

La Adquisición de Conocimientos consiste en la recolección de la información necesaria para construir un Sistema Basado en el Conocimiento a partir de cualquier posible fuente. Esta información puede estar constituida por datos, noticias, conocimientos humanos, etcétera.

La Adquisición de Conocimientos no debe ser considerada como una etapa dentro de una metodología para la construcción de un Sistema Basado en el Conocimiento sino un proceso paralelo a todas las fases de desarrollo de uno de estos sistemas, ya que cada etapa necesita determinada información, lo que provoca que la recolección de ésta no se haga en un único paso aislado sino en cada una de las etapas. El papel de la Adquisición del Conocimiento en las primeras fases (definición del problema, conceptualización, ...) es fundamental, mientras que en las últimas etapas del desarrollo (implementación, evaluación, mantenimiento), la dedicación a la Adquisición de Conocimiento es mucho menor.

Es posible que la tarea más importante para el desarrollo de un Sistema Basado en el Conocimiento sea la Adquisición del Conocimiento. Pero paradójicamente, éste es un campo experimental más que una tecnología, y salvo en el caso de la inducción y aprendizaje automático, la Inteligencia Artificial no aporta métodos completos que solucionen o automaticen esta tarea, sólo técnicas para abordar problemas parciales. Por consiguiente, la Adquisición del Conocimiento resulta en la actualidad una labor artesanal, propia para cada caso y dependiente de las personas concretas que estén involucradas en ella. Todo esto provoca que uno de los principales *cuellos de botella* en el desarrollo de un Sistema Basado en el Conocimiento sea el adquirir los conocimientos necesarios para poder construir sistemas eficientes. La información necesaria puede presentarse de múltiples formas, aunque conviene, en lo que respecta a los Sistemas Basados en el Conocimiento, considerar especialmente una serie de fuentes:

- *Libros y manuales.* Conocimientos básicos, específicos y públicos del dominio y del problema.
- *Documentación formal.* Documentos que contienen políticas, procedimientos, estándares, normas, regulaciones, leyes, etcétera, de un dominio. Este conocimiento también es de carácter público.
- *Documentación informal.* Notas, manuscritos, ayudas de trabajo, etcétera, proporcionan frecuentemente conocimiento heurístico para la resolución de problemas. Aunque a veces esta documentación es confidencial proporciona conocimientos semipúblicos.
- *Registros internos.* Registros de casos que se presentan, en forma de fichas de clientes, pacientes, estudios, estadísticas, etcétera. Pueden estar en forma escrita o, cada vez más, en forma magnética (bases de datos). Además de para la validación y evaluación de los Sistemas Basados en el Conocimiento, esta información debe ser útil para la generación del propio conocimiento del sistema, y éste es uno de los objetivos de este trabajo.
- *Presentaciones.* Material usado para la formación, impartida o recibida. Tiene la ventaja de contener conocimientos expuestos de una forma muy clara.
- *Publicaciones.* Revistas especializadas, actas de congresos, etcétera.
- *Investigación.* Resultados de las investigaciones que se estén llevando a cabo, en forma de datos empíricos, informes, resultados estadísticos, etcétera.
- *Visitas.* El Ingeniero de Conocimiento se desplaza a los centros de trabajo del experto y los usuarios, para observar "in situ" el *modus operandi*.

- *Conocimiento humano.* Además de las entrevistas con los expertos, resulta imprescindible la interacción con los directivos y usuarios. Los directivos pueden aportar objetivos del proyecto, alcance del sistema, contexto donde irá instalado, etcétera. Los usuarios deben dar claves de interfaces, necesidades, requisitos, etcétera.

El Ingeniero de Conocimiento debe controlar constantemente qué información necesita, con qué profundidad, sobre qué temas, que técnica debe utilizar para adquirir ese conocimiento y otros factores. Muchas veces resulta tentador improvisar, lo que frecuentemente provoca resultados negativos y falta de rigor. Por estas razones, el método que se presenta en este trabajo, contempla básicamente tres grandes bloques:

1. Evaluación de viabilidad, definición del problema, primeras reuniones.
2. Extracción de conocimientos de la documentación (incluidas bases de datos y demás fuentes documentales anteriormente mencionadas).
3. Educación de conocimientos del experto, directivos y usuarios.

De una forma más precisa, los objetivos de la Adquisición de Conocimiento son los siguientes:

1. Comprensión general de la tarea y estructura funcional del Sistema Basado en el Conocimiento.
2. Proceso de razonamiento del experto y pasos en la resolución de problemas.
3. Datos necesarios para resolver un problema determinado, con los valores que pueden tomar.
4. Desarrollo de un modelo conceptual.

Además de los métodos usuales de Adquisición de Conocimiento: Métodos automáticos, a partir de ejemplos, inducción. Análisis estructural de textos. Entrevistas (Abiertas, estructuradas). Observación de tareas habituales. Clasificación de conceptos. Cuestionarios. Análisis de protocolos. Emparrillados (*Repertory Grids*). Técnicas para educación en grupo. Método Delphi, se usa *Data Mining* para manipular grandes volúmenes de datos, con el fin de proporcionar información útil y conocimiento a partir de los mismos.

2.1. El proceso KDD.

Fayyad y sus colaboradores definen el proceso KDD como “*El proceso no trivial de identificar válidos, nuevos, potencialmente útiles y comprensibles patrones en datos*”, utilizando técnicas multidisciplinares y viendo como actúan juntas. El término *proceso* implica que hay varios pasos, como la preparación de datos o la búsqueda de patrones. El término *patrón* (en este trabajo es el denominado *prototipo de datos*) se refiere a un subconjunto de datos, junto con una descripción y un modelo aplicable al mismo. Los *prototipos de datos* descubiertos deben ser válidos para nuevos datos con cierto grado de certeza. Se necesita que estos patrones sean nuevos, al menos para el sistema y preferiblemente para el usuario y potencialmente útiles. Por último, estos patrones deben ser comprensibles, todo ello si no es inmediatamente, lo debe cumplir después de un postprocesado. Esta definición implica que deben definirse medidas de la bondad de los prototipos de datos, en muchos casos es posible definir medidas de certeza (capacidad de clasificación de nuevos datos) o utilidad (calidad de las predicciones en base a estos prototipos de datos).

El proceso KDD descrito por Fayyad y sus colaboradores, interactivo e iterativo, es el que refleja la fig. 80:

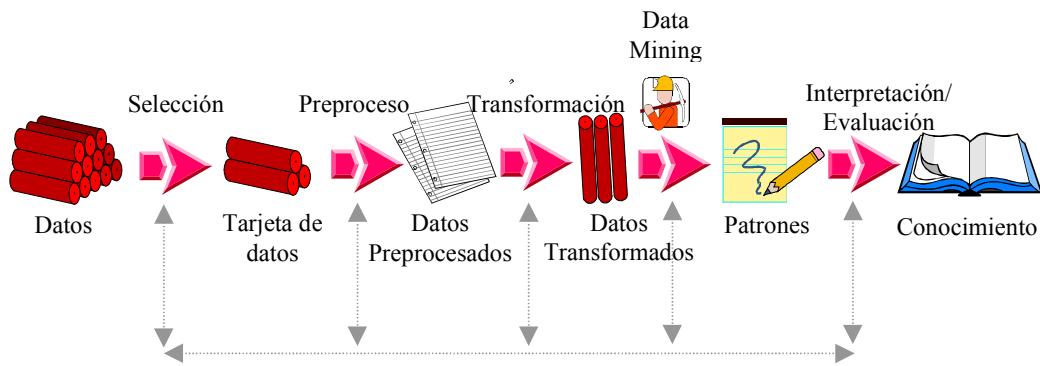


Figura 80.

- **Selección:** Aplicando el conocimiento del dominio y el conocimiento relevante a priori, teniendo en cuenta los objetivos del proceso global de DCPB se crea una tarjeta de datos que incluirá conjuntos seleccionados de datos o subconjuntos de variables relevantes o ejemplos.
- **Preproceso:** Limpieza de datos, eliminación de ruidos, manejo de campos vacíos, datos perdidos, valores desconocidos o por defecto, evolución de datos. Se aplican técnicas estándar de bases de datos.
- **Transformación:** Reducción del número de variables. Localización de formas útiles para expresar los datos dependiendo del uso posterior que se les va a dar y de los objetivos del sistema. Se usa el conocimiento experto y técnicas de transformación e informes en bases de datos.
- **Data Mining:** Elección de los algoritmos de Data Mining. Decisiones acerca del modelo que se deriva del algoritmo de Data Mining elegido (clasificación, resumen de datos, predicción). Búsqueda de patrones de interés, en cuanto a clasificación, reglas o árboles, regresión, clasificación, dependencia, heurísticas, incertidumbre.

2.1.1. Algoritmos de Data Mining.

Dentro del proceso *KDD* descrito, adquiere especial relevancia el paso de *Data Mining* para determinar los patrones de los datos observados. La elección de los modelos a utilizar tiene una componente fundamental de conocimiento de los expertos, supervisado por el Ingeniero de Conocimiento. En la literatura, véase por ejemplo [Berry, 96], se describe un gran número de algoritmos y técnicas de Data Mining. En el método propuesto en este trabajo se utilizan fundamentalmente los provenientes de la *lógica borrosa*, la estadística y la gestión de bases de datos en lo que concierne al proceso *DCPB*; Para la generación y manipulación de los prototipos conceptuales, se usan técnicas de *reconocimiento de patrones* y *lógica borrosa* principalmente. En particular, los algoritmos de *Data Mining* consisten en una mezcla específica de tres componentes [Fayyad, 96b]:

2.1.1.1. El modelo.

Un modelo contiene parámetros que son determinados a partir de los datos. Dos factores relevantes:

La función.

En la práctica del Data Mining, se utilizan funciones de *Regresión* (comparan datos con valores reales de las variables establecidos mediante predicción), *Dependencia* (describen dependencias significativas entre variables), *Análisis de relaciones* entre campos de las bases de datos, pero las más relevantes para el modelo presentado en este trabajo, cuyo objetivo es la predicción (como evaluación de una situación mediante la comparación de casos reales con situaciones prototípicas), son las siguientes:

- Funciones de *Clasificación*: Categorizan un registro dentro de una de varias clases predefinidas.
- Funciones de *Clustering*: Categorizan un registro dentro de una de varias clases (*clusters*), pero a diferencia de la clasificación, las clases las determinan los propios datos, mediante agrupamientos naturales basados en medidas de afinidad, similaridad o probabilidad.
- Funciones de *Resumen*: Generan una descripción compacta de un subconjunto de datos. Se puede usar un simple ejemplo como media y las desviaciones estándar para todos los campos.
- Funciones de *Análisis de secuencias*: Modelizan patrones secuenciales, series temporales. El objetivo es generar la secuencia de estados del proceso que se trata de modelizar.

La representación.

Se utilizan modelos clásicos, como árboles de decisión, reglas, modelos lineales y no lineales, métodos basados en ejemplos (razonamiento basado en casos), redes bayesianas, modelos borrosos, redes neuronales, algoritmos genéticos, etcétera. El modelo de representación determinará la flexibilidad de representación de los datos y interpretación del modelo desde el punto de vista humano. Normalmente, cuanto más complejo sea el modelo de representación, es posible que maneje mejor los datos, pero hará que sea mucho más difícil entenderlos. Aunque a nivel de investigación se tiende a utilizar modelos complejos, en aplicaciones reales se utilizan mayoritariamente modelos simples debido a su robustez e interpretabilidad.

2.1.1.2. El criterio de preferencia.

Es aquel que permite seleccionar un modelo o conjunto de parámetros en función de unos datos determinados. Es, en cierto modo, una medida de bondad de la relación entre el modelo y los datos. Normalmente este criterio es explícito y cuantitativo en el algoritmo de búsqueda (por ejemplo, el criterio de encontrar los parámetros que maximicen la probabilidad de algunos datos observados). También suele haber un criterio subjetivo (por parte del Ingeniero de Conocimiento) implícito, sobre qué modelos se ha de tener en cuenta inicialmente.

2.1.1.3. El algoritmo de búsqueda.

Es necesario especificar un algoritmo para encontrar modelos particulares y parámetros e incluso criterios de preferencia. En los de búsqueda de parámetros es frecuente que, dado un modelo, el problema de encontrar los mejores parámetros se reduzca a un problema de optimización. En *Data Mining* se suelen utilizar técnicas de optimización sencillas (por ejemplo, el *descenso del gradiente*), que presentan los problemas de máximos y mínimos locales y otros muy estudiados en el ámbito de la Inteligencia Artificial.

2.2. Conceptos y notación para los algoritmos de aprendizaje automático e inducción.

2.2.1. Algoritmos de clustering jerarquizado.

Repertory Grids o Emparrillados. En el año 1955, el psicólogo Kelly desarrolló un modelo de pensamiento humano que denominó “Personal Construct Theory”. Según esta teoría, cada persona es como un científico con su propio modelo personal del mundo que le rodea. Esta persona clasifica y jerarquiza su mundo, desarrollando una serie de teorías sobre él. Basándose en estas teorías es posible anticiparse a la reacción que puede tener él frente a una serie de estímulos.

Este modelo reúne las cualidades de ser como la visión que tiene un experto en su trabajo, es una descripción del contenido y uso de su conocimiento. Esta teoría ha sido criticada porque se ajusta muy particularmente a la visión de una persona. Así, varios expertos pueden tener diferentes puntos de vista que dan lugar a diferentes respuestas. Esto precisamente puede ser explotado a

favor de la teoría. Hay evidencia (estudios realizados por Newell y Simon en 1972) de que el camino en que los expertos contemplan los problemas es el que los distingue de los inexpertos o recién incorporados.

El camino seguido para la clasificación refleja y constituye parte de la experiencia del experto. Puede ayudar a diferenciar las características superficiales frente a las verdaderas leyes que rigen los procesos. Por ejemplo un inexperto podría afirmar: "Mi programa no lee todos los datos del fichero", donde un experto diría "La estructura del bucle de control del programa está mal". Basada en esta teoría, surgen las denominadas "Mallas de repertorios" o "Emparrillados" (*Repertory Grids*).

2.2.1.1. Emparrillados.

Un emparrillado consta de **construcciones** (*constructs*) y de **elementos**, equivalentes a los atributos y sus ejemplos. Una construcción es una característica bipolar o dicotómica que posee cada elemento en algún grado. Sirve para distinguir unos elementos de otros o verificar su similitud. Por ejemplo serían construcciones grande-pequeño o muchos días-pocos días. Un elemento es una entidad abstracta o concreta que es un caso importante desde la perspectiva del experto. El proceso de adquisición de conocimiento consiste en obtener (solicitar del experto o por otras técnicas) todos los elementos y construcciones relativas al dominio de conocimiento a explorar.

Las construcciones contemplan valores graduales (habitualmente discretos) entre sus polos, como podrían ser las etiquetas lingüísticas de un *term-set* borroso, por ejemplo {Muy pesado, pesado, peso medio, ligero, muy ligero}. La escala puede ser simbólica o numérica. No puede variar en toda la malla el número de escalones (normalmente 1 a 3 ó 1 a 5), aunque las etiquetas utilizadas para cada construcción sean diferentes. Con los elementos y las construcciones se crea una red, en la que habitualmente las construcciones se disponen en filas y los elementos en columnas (dependiendo del numero de elementos o construcciones).

Por ejemplo, supóngase esta malla, con una escala de 1 a 5, las columnas representan los elementos y las filas las construcciones:

	Miguel	Juan	Ana	Jorge	María	
	1	← →		5		
Inteligente	1	1	4	5	3	Torpe
Tímido	1	1	5	4	2	Extrovertido
Inexperto	1	2	3	5	4	Experto
Motivado	1	1	4	5	2	No motivado
Informal	3	2	2	5	1	Formal

Tabla 76: Ejemplo de parrilla.

Análisis de la malla.

Puede ser realizado por varios métodos:

- Análisis de *clusters* o grupos.
- Aplicación de la *teoría de la confirmación*.
- Análisis factorial.
- Análisis basado en lógica borrosa.

El primer tipo consiste en encontrar grupos homogéneos de elementos o de construcciones. Para hacer este análisis es necesario definir una medida de similitud, como una función que asocia un valor numérico a cualquier par de vectores para ver su similitud. Con esta función se construye una

matriz cuadrada de proximidad, donde cada valor mide la similitud entre el elemento fila y columna que se presenta.

Esta matriz será la entrada al análisis de *clusters*. Aunque aquí también se pueden utilizar varias técnicas, se usará la denominada clustering jerárquico.

Un *cluster* es una subdivisión de un conjunto de objetos E en varios subconjuntos E_i , que cubren E de forma que:

$$E = \bigcup_i E_i \text{ y } E_i \cap E_j = \emptyset \text{ para } i \neq j$$

Las funciones de similitud utilizadas, en las que cuanto más distantes sean los vectores más grande será d, son las siguientes:

Sean u y w vectores:

$$u = (x_1, x_2, \dots, x_n)$$

$$w = (y_1, y_2, \dots, y_n)$$

Distancia entre vectores u y w , $d(u, w)$:

$$\text{Puede ser } d(u, w) = \sum_i |x_i - y_i| \quad \text{o} \quad d(u, w) = \sum_i |x_i - y_i|^2$$

Ejemplo:

Sea la Malla de Repertorio dada en la tabla 77, donde 01 ... 06 son elementos y A1 ... A6 construcciones:

	01	02	03	04	05	06	
A1	1	5	3	1	5	1	A1
A2	3	5	5	3	5	1	A2
A3	3	1	5	5	3	3	A3
A4	3	1	1	3	3	5	A4
A5	3	5	3	3	5	1	A5
A6	1	5	3	1	3	1	A6

Tabla 77: Malla de repertorio para el ejemplo.

Para realizar un Análisis de clusters sobre elementos, se construye una matriz de proximidad, en base a las distancias calculadas según la función descrita anteriormente (la primera):

$$d(01, 02):$$

$$\begin{array}{ccccccc} 1 & 3 & 3 & 3 & 3 & 1 \\ & \underline{5} & \underline{1} & \underline{1} & \underline{5} & \underline{5} \\ 4+2+2+2+2+4 & = 16 \end{array}$$

$$d(03, 04):$$

$$\begin{array}{ccccccc} 3 & 5 & 5 & 1 & 3 & 3 \\ & \underline{1} & \underline{3} & \underline{5} & \underline{3} & \underline{3} \\ 2+2+0+2+0+2 & = 8 \end{array}$$

$$d(02, 06):$$

$$\begin{array}{ccccccc} 5 & 5 & 1 & 1 & 5 & 5 \\ & \underline{1} & \underline{1} & \underline{3} & \underline{5} & \underline{1} \\ 4+4+2+4+4+4 & = 22 \text{ (muy diferente).} \end{array}$$

... Así se calcularían todas las distancias, dando lugar a lo que se denomina “malla de repertorio reducida”, tabla 78.

	01	02	03	04	05	06
01	--	16	10	2	10	6
02		--	10	18	6	22
03			--	8	8	16
04				--	12	8
05					--	16
06						--

Tabla 78: Malla de repertorio reducida.

Transformando las distancias a porcentajes de proximidad: $s_{ij}(\%) = 100 - \frac{100 * d(i, j)}{D * n}$

n = número de construcciones.

D = diferencia máxima posible para una construcción.

(Cuando $d(i, j) = D \cdot n$, máxima diferencia. En el ejemplo $D = 4$, $n = 6$), se obtiene la tabla reducida a porcentaje (tabla 79):

	01	02	03	04	05	06
01	--	33	58	92	58	75
02		--	58	25	75	8
03			--	67	67	33
04				--	50	67
05					--	33
06						--

Tabla 79: Tabla reducida a porcentajes de proximidad.

En caso de necesitarse, una posible optimización del método consistiría en utilizar matrices borrosas y conseguir matrices borrosas equivalentes óptimas [Fu, 98].

Los elementos con mayor porcentaje de proximidad son el 01 y el 04, con un 92%, por lo que se pueden agrupar como un solo objeto (01, 04). Ahora al calcular la similitud de este objeto compuesto, se elige el mínimo de las similitudes con sus componentes. Por ejemplo (01, 04) con 02 será:

$$\min [(01, 02), (04, 02)] = \min (33, 25) = 25$$

ó (01,04) con 05: $\min [(01, 05), (04, 05)] = \min (58, 50) = 50$

Así se pueden seguir desarrollando las tablas:

	(01, 04)	02	03	05	06
(01, 04)	--	25	58	50	67
02		--	58	75	8
03			--	67	33
05				--	33
06					--

Tabla 80: Mayor similitud (01, 04).

Ahora la mayor similitud es entre 02 y 05, un 75%:

	(01, 04)	(02, 05)	03	06
(01, 04)	--	25	58	67
(02, 05)		--	58	8
03			--	33
06				--

Tabla 81: Mayor similitud (02, 05).

Los más similares son ahora (01, 04) y 06, con un 67%:

	[(01, 04), 06]	(02, 05)	03
[(01, 04), 06]	--	8	33
(02, 05)		--	58
03			--

Tabla 82: Mayor similitud [(01, 04), 06].

Entre (02, 05) y 03 se da la mayor similitud, un 58%:

	[(01, 04), 06]	[(02, 05), 03]
[(01, 04), 06]	--	8
[(02, 05), 03]		--

Tabla 83: Mayor similitud [(02, 05), 03].

Por último se observa que [(01, 04), 06] y [(02, 05), 03] son poco similares (8%). Con esta información se puede construir un “Dendograma de clustering jerárquico”, que permite ver la relación entre elementos, separándolos en clases no definidas *a priori*:

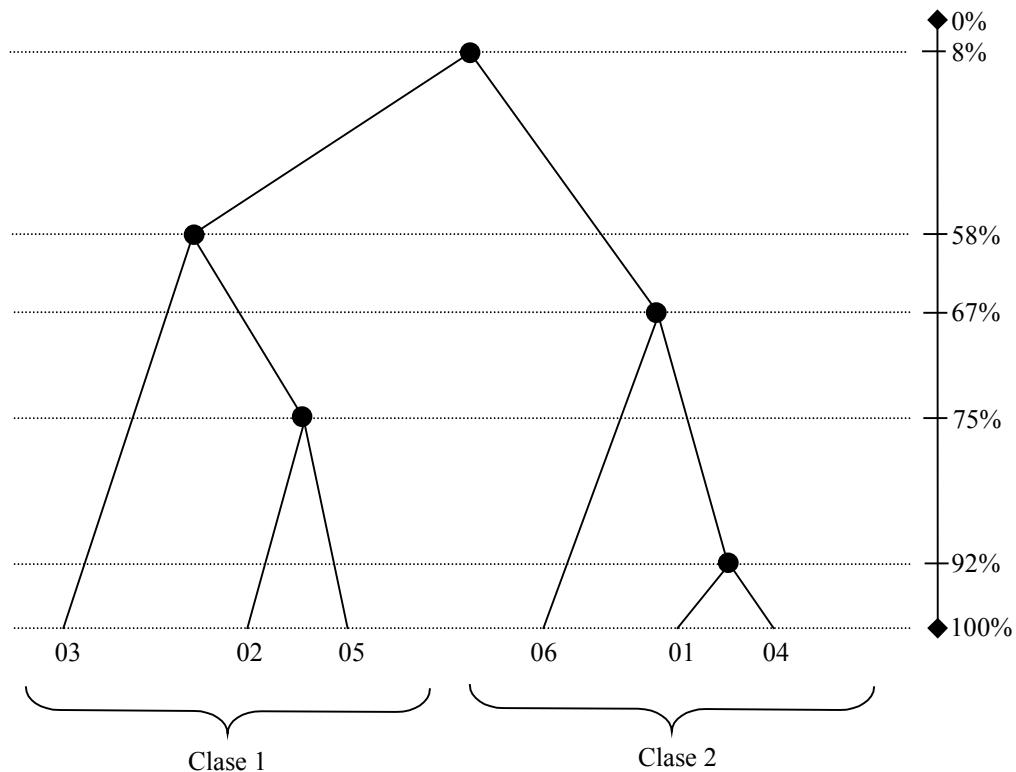


Figura 81: Dendrograma de *clustering* jerárquico.

También es posible realizar una **Comparación de construcciones**. Es algo más compleja que la de elementos porque hay que comparar las construcciones y las construcciones invertidas u opuestas. Por ejemplo, si comparamos características físicas de personas a través de las construcciones bajo/alto, mucho peso/poco peso podría no haber similitud entre ellas, pero si invertimos una lo más seguro es que haya una mayor semejanza (bajo relacionado con poco peso y alto con mucho peso).

Una construcción invertida X' se obtiene de la construcción original X , sustituyendo los valores en X por sus opuestos en la escala de valoración. Por ejemplo, si la escala es de 1 a 3, el opuesto al 1 es el 3, el del 2 el 2 y el del 3 el 1. Un caso podría ser:

	E1	E2	E3	E4
C1	3	1	3	2
C1'	1	3	1	2

La matriz para la comparación de construcciones tiene por encima de la diagonal los valores iniciales (por ejemplo C1, C2) y por debajo se considera la inversión de uno de ellos (para el caso C2, C1'). El proceso posterior es el mismo que en el análisis de *clusters* sobre elementos.

2.2.2. Algoritmos de inducción.

Se presentan algunos conceptos y notación necesarios para la aplicación posterior en el proceso de *Data Mining* sobre la base de datos de incendios forestales. Los algoritmos usados son ID3 y CART, aunque también se contempló la posibilidad de utilizar C4.5 [Quinlan, 88], pero esta extensión de ID3 no aportaba nada significativo para el uso que se pretende en esta aplicación.

2.2.2.1. Conjunto de entrenamiento.

El aprendizaje inductivo se realiza extrayendo conocimientos a partir de un conjunto de ejemplos externos. Son ejemplos de decisiones tomadas previamente por un experto e incluyen rasgos que describen el ejemplo así como la decisión tomada (lo que se definirá como aprendizaje *supervisado*). Sea un ejemplo de clasificación de setas entre venenosas y comestibles. Aquí los ejemplos constarían de los rasgos de las setas: color, tamaño, ... Así como la clasificación por expertos en venenosas y comestibles. El conjunto de ejemplos es el conjunto de entrenamiento, aunque a veces se usa parte como test y parte como conjunto de entrenamiento.

2.2.2.2. Rasgos o Atributos.

Son difíciles de extraer automáticamente, siendo mucho más sencillo hacerlo de forma manual. Son las características de los ejemplos del dominio. Básicamente pueden ser de dos tipos:

- **Simbólicos:** Aquellos individuos cuyos valores no permiten establecer ningún tipo de orden entre ellos.
- **Ordenados:** tiene sentido establecer una relación de orden entre los distintos valores que puedan tomar (edad de una persona...).

Si los Atributos son discretos (tienen un número fijo o limitado de valores):

- **Binarios:** indican presencia o ausencia de atributos.
- **Multivaluados.**

Cuando se trata de magnitudes continuas o reales se suelen representar estos atributos usando varios intervalos en el rango de posibles valores.

Datos con **ruido o incompletos**: Los datos con ruido se generan debido a la corrupción de la base de ejemplos y los incompletos por una falta de información.

Tipos de aprendizaje (supervisado frente a no supervisado):

- **Aprendizaje supervisado:** Cada ejemplo consta de un conjunto de condiciones iniciales y un conjunto de acciones resultantes o decisiones.
- **Aprendizaje no supervisado:** Se dispone sólo de los ejemplos de entrada. El objetivo es comprender la estructura estadística o modelo que ofrecen los datos.

2.2.2.3. Análisis del conjunto de ejemplos.

Si el número de ejemplos disponibles no es significativo se pueden obtener falsas conclusiones. El número de ejemplos debe cubrir razonablemente la región de interés del espacio de entradas (es decir, debe quedar el menor número de valores de los atributos sin aparecer en algún ejemplo). Si hay demasiados atributos para caracterizar el ejemplo entonces el proceso de aprendizaje es más largo y el número de ejemplos necesario mayor para cubrir el espacio de entradas. Cada ejemplo ha de estar representado mediante una serie de atributos que permitan diferenciarlo del resto.

El **Sobreaprendizaje** es un problema que consiste en aprender demasiado sin tener capacidad de generalización. Un modelo generaliza bien cuando aprende de modo global, cuando memoriza lo significativo, desecharlo lo que se debe únicamente al azar. Se controla buscando subconjuntos de ejemplos significativos, tratando de eliminar el *ruido* y las redundancias en los ejemplos. La validación de estos modelos se realiza mediante conjuntos de ejemplos-test.

2.2.2.4. Algoritmo de aprendizaje.

Aprende a partir de un conjunto de ejemplos y produce conocimiento simbólico. Los tipos de estrategias que se usan son: De lo general a lo específico y de lo específico a lo general.

Dos tipos de algoritmos de aprendizaje: Algoritmos que dividen el espacio de entrada y algoritmos que realizan una proyección en el espacio de salida. Sea un espacio de entrada y uno de salida. Un punto del espacio de entrada representa valores específicos de los atributos. Un punto en el espacio de salida representa una clase en la que clasificar.

Un algoritmo de partición del espacio de entradas divide progresivamente dicho espacio seleccionando atributos en una secuencia que parte este espacio. Típicamente se va escogiendo un atributo a la vez. En un momento dado no es necesario subdividir más el espacio de entrada y en este momento se asigna esta partición a una clase dada de salida.

Se necesita en este método:

- Una forma de generar la siguiente partición (implica la selección de un atributo para subdividir)
- Un criterio de parada de subdivisiones.
- Una regla para asignar la partición a una salida.

Más de una partición puede asignarse a una salida dada.

Un algoritmo de proyección del espacio de salida selecciona una salida en concreto que es un punto del espacio de salida, y determina la región del espacio de entrada que le corresponde. Esto se repite para cada posible salida.

Aquí se necesitaría:

- Un método para relacionar una salida dada a puntos en el espacio de entrada.
- Un criterio de selección de puntos por esta relación.
- Un criterio de identificación de regiones frontera de interés en el espacio de entrada.

2.2.3. Algoritmos de inducción usando árboles de decisión.

Hay 3 familias importantes:

- CLS (Concept Learning System, Hunt 1966), ID3 (Interactive Dichotomic 3, Quinlan, 1986), ACLS.
- CART (Classification and Regression Trees, Breiman 1984)
- Uso de criterios estadísticos.

Rasgos comunes de estos sistemas:

- Uso de una estrategia de arriba-abajo.
- Uso de una estrategia de divide y vencerás.

Diferencias más notables: criterios de selección, en cómo se divide el conjunto de ejemplos en cada paso.

2.2.3.1. Definición de árbol de decisión

Es un grafo dirigido que consta de nodos y arcos dirigidos. Los nodos corresponden a una pregunta o test. Un árbol de decisión se puede ver como un modelo de aprendizaje inductivo a partir de un conjunto de ejemplos o conjunto de entrenamiento, donde la información extraída del mismo queda organizada jerárquicamente en forma de árbol. Los árboles de decisión están formados por nodos y arcos que unen nodos. Los nodos tienen asociado un separador que típicamente es una pregunta acerca de la existencia o no de un atributo en el caso a clasificar.

2.2.3.2. Aprendizaje de árboles de decisión.

El esquema de los sistemas a utilizar es el siguiente:

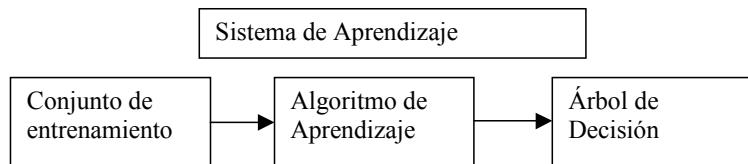


Figura 82: Esquema de los sistemas de aprendizaje por inducción.

El algoritmo de aprendizaje usa el criterio de selección de rasgos para seleccionar un atributo "separador" para dividir el conjunto de entrenamiento. El conjunto de entrenamiento es subdividido progresivamente usando los separadores obtenidos a través del criterio de selección de rasgos. Una hoja en el árbol corresponde a una sola clase. La clase se asigna cuando todos los ejemplos en la subdivisión particular tienen la misma clase.

2.2.4. CLS.

CLS (*Concept Learning System*) [Hunt, 66] es el patriarca de la familia de algoritmos que obtienen árboles de decisión. CLS construye un árbol de decisión que intenta minimizar el coste de clasificar un ejemplo. Este coste tiene dos componentes:

- Clasificar el ejemplo correctamente.
- Clasificar el ejemplo incorrectamente.

CLS usa una estrategia de mirada-haciaadelante similar a MINIMAX, que es una estrategia que explora el espacio de posibles árboles de decisión hasta un nivel de profundidad prefijado, elige una acción para minimizar el coste en este espacio limitado y entonces se mueve un nivel abajo en el árbol. En su forma completa, el cálculo requerido crece exponencialmente con la profundidad de la mirada hacia adelante empleada.

La estrategia del algoritmo es la siguiente:

Sea un conjunto de entrenamiento S que consta de varios ejemplos, cada uno perteneciente a una de las clases $C_1, C_2, C_3, \dots, C_n$. El algoritmo tiene los dos pasos siguientes:

1. Si todos los ejemplos en el conjunto S pertenecen a la misma clase C_i , el árbol de decisión para el conjunto S consta de una hoja etiquetada con esa clase.
2. Sea T algún test con resultados posibles O_1, O_2, \dots, O_n . Cada ejemplo en S tiene un resultado para T tal que el test parte S en subconjuntos S_1, S_2, \dots, S_n donde cada ejemplo en S_i tiene resultado O_i para T .

T resulta ser raíz del árbol de decisión y para cada resultado, se construye un subárbol de decisión aplicando recursivamente el algoritmo.

2.2.5. Grupo de algoritmos de ID3.

ID3 (*Interactive Dichotomizer 3*) [Quinlan, 86] es uno de los programas desarrollados a partir de CLS (*Concept Learning System*) [Hunt, 66]. Su funcionamiento consiste en que un conjunto de ejemplos es presentado al algoritmo de aprendizaje. A partir de ellos el algoritmo desarrolla un árbol de decisión de arriba a abajo, guiado por la frecuencia de la información de los ejemplos. El orden de presentación de los ejemplos no es importante

El algoritmo de aprendizaje requiere el examen y reexamen de todos los ejemplos en muchas etapas durante el aprendizaje. Es un enfoque "dirigido por datos", usando un método de inducción de arriba a abajo para producir su árbol de decisión. La estructura básica del ID3 es iterativa.

El objetivo del método es producir un árbol que clasifique correctamente todos los ejemplos en un subconjunto de elementos del conjunto de entrenamiento, y consiste en lo siguiente:

1. Realiza una partición recursiva del espacio de entradas, asociando a cada una de las zonas generadas una clase.
2. El árbol se genera dividiendo sucesivamente el conjunto de aprendizaje en subconjuntos de ejemplos cada vez más pequeños hasta conseguir conjuntos suficientemente puros.
3. Un conjunto es bastante puro cuando casi todos los ejemplos de él pertenecen a una sola clase (impuro es lo contrario).

Los elementos clave del proceso recursivo ID3 (fig. 83), son los siguientes:

- Modo de selección del mejor separador
- Criterio para decidir expansión o no

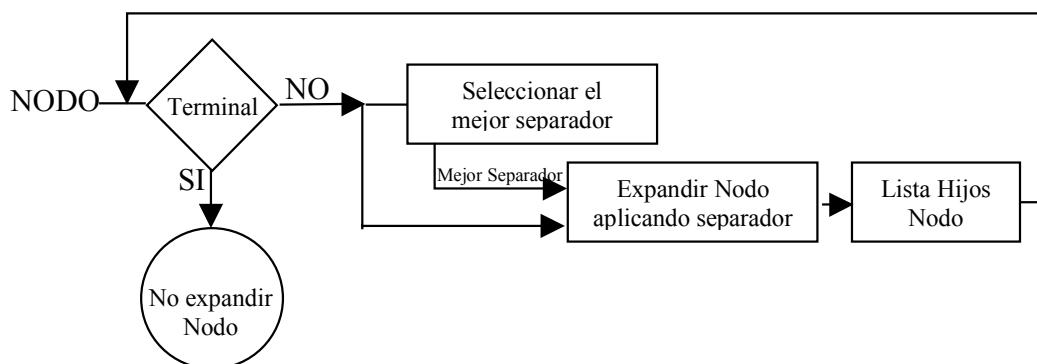


Figura 83: Proceso recursivo ID3.

2.2.5.1. Selección del mejor separador.

El separador suele ser un solo atributo ya que tiene significado físico y da como resultado un árbol sencillo de interpretar. Dos tipos de separadores:

- Simbólico: el separador será una pregunta del tipo ¿Cuál es el valor del atributo X?. Puede haber más de 2 hijos.
- Ordenado: ¿es el valor del atributo menor que un umbral determinado?.

Para seleccionar el mejor separador hay que utilizar una medida para distinguir entre posibles candidatos. La medida que se utiliza se basa en la entropía o cantidad de información definida en la teoría de la información, véase mejor con un ejemplo:

Sean :

S el conjunto de entrenamiento
x, y, z, números de ejemplos de las clases X,Y,Z.

Hipótesis 1:

Cualquier árbol de decisión correcto para S clasificará los ejemplos en la misma proporción con que ellos están representados en S. Un ejemplo arbitrario pertenecerá a las clases X,Y,Z con probabilidad:

$$x / (x + y + z) ; y / (x + y + z) ; z / (x + y + z), \text{ respectivamente.}$$

Hipótesis 2:

Cuándo un árbol de decisión es usado para clasificar un ejemplo, devuelve una clase. Un árbol de decisión puede considerarse como una fuente de mensajes X, Y o Z. La información expresada requerida para generar este mensaje vendrá dada por

$$I(X, Y, Z) = -\frac{x}{x+y+z} \log_2 \left(\frac{x}{x+y+z} \right) - \frac{y}{x+y+z} \log_2 \left(\frac{y}{x+y+z} \right) - \frac{z}{x+y+z} \log_2 \left(\frac{z}{x+y+z} \right)$$

Si se escoge un atributo A como separador su contenido informativo será :

$$E(A) = \sum \frac{x_i + y_i + z_i}{x+y+z} I(x_i, y_i, z_i)$$

x_i, y_i, z_i , son números de ejemplos de las clases X, Y, Z con valor A_i del atributo A (multiatributo).

La información generada por el atributo separador A o bien la disminución de la entropía vendrá dada por:

$$\Delta I(A) = I(X, Y, Z) - E(A)$$

Compararemos esta medida para los diferentes atributos y decidiremos cuál va a ser el mejor separador.

Generalizando,

- Entropía de un nodo N

$$I(N) = - \sum_{i=1, NClases} P(N, C_i) \log_2 [P(N, C_i)]$$

Nclases : número de clases diferentes en el nodo.

$P(N, C_i)$: proporción de ejemplos en el nodo pertenecientes a la clase C_i .

- La Selección del mejor separador se realiza eligiendo el que produce una disminución mayor en la impureza del nodo analizado

$$\Delta I(N, S) = I(N) - \sum_{i=1, NHijos} P(Nh_i) I(Nh_i)$$

$\Delta I(N, S)$: decremento de impureza del nodo N producido al aplicar el separador S.

$I(N)$: impureza del nodo.

N: número de hijos generados al aplicar S.

$P(Nh_i)$: proporción de ejemplos del nodo N que quedan en el nodo hijo i.

2.2.5.2. *Criterios para detener la expansión de un nodo*

Existen varios criterios para decidir si el aumento de la capacidad de clasificación del árbol al expandir un nodo es suficientemente significativo como para compensar el incremento de complejidad del árbol.

Normalmente se compara la impureza del nodo en estudio con un valor mínimo H_{min} . Si la impureza del árbol es menor, el nodo no se expande.

2.2.6. *La familia CART (Classification and Regression Trees).*

CART (*Classification and Regression Trees*) [Breiman, 84] es el principal algoritmo del segundo grupo de programas desarrollados a partir del algoritmo de Hunt CLS. CART construye un árbol de decisión dividiendo repetidamente el conjunto de entrenamiento en subconjuntos.

Su principal objetivo es encontrar un camino sistemático para predecir la clase a la que un objeto en particular pertenece. Aparte de producir un exacto clasificador, intenta también descubrir la

estructura predictiva del problema. Una característica importante de CART es su algoritmo de poda.

2.2.6.1. Construcción del árbol.

La construcción del árbol depende de:

- La selección de los separadores.
- La decisión para continuar dividiendo o no un nodo
- La asignación de una clase a un nodo no dividido (no expandido).

Un separador divide el conjunto de entrenamiento en dos subconjuntos. En algún aspecto es similar a ID3, pero se diferencia de éste en que en CART sólo admite divisiones binarias.

Cuando se selecciona un separador, el objetivo es optimizar la homogeneidad de los subconjuntos resultantes, o sea, CART intenta seleccionar un divisor tal que los subconjuntos resultantes tengan ejemplos de una sola clase.

La bondad de un separador se tiene que definir en base a la medida de pureza o impureza. Un árbol crece de la siguiente manera: Inicialmente una búsqueda es hecha a través de todos los posibles separadores candidatos para encontrar el separador con el mayor decremento de la medida de impureza. La expansión de un nodo se para cuando no hay descenso significativo de impureza. La clase asignada a este nodo no expandido se determina por reglas de pluralidad.

De los posibles separadores se escoge aquel que produzca la mayor disminución de impureza en los conjuntos. Existen dos criterios principales para determinar el mejor separador: El criterio de Gini y el criterio de Twoing.

2.2.6.2. Criterio de Gini.

Da un índice del grado de diversidad de información en el nodo (cuanto menor sea mejor) . Es el criterio más usado. La forma del índice de Gini es:

t: nodo que está siendo considerado.

$$I(t) = \sum_{j \neq i} p(j | t) p(i | t)$$

p(i|t), p(j|t) probabilidades de que el nodo t esté en la clase i ó j.

Estas probabilidades se estiman a veces como la proporción de las clases en el nodo. Otras veces pueden ser estimadas del conjunto de entrenamiento o suministradas por el analista.

El índice de Gini es simple y puede ser rápidamente calculado. El único problema con el criterio de Gini es que la selección de atributos está desviada a favor de aquellos que tengan más valores.

2.2.6.3. Criterio de Twoing.

Es un método alternativo para la medida de la impureza. No da una medida global de impureza como hace Gini sino que da una idea aproximada de la impureza para construir el árbol. La estrategia seguida es reducir el problema a un problema de dos clases. En cada nodo, las clases son separadas en dos superclases denominadas clase 1 y clase 2. Se tiene un problema con dos clases.

Los separadores son estratégicos, y el usuario es informado de las similaridades de clases. Arriba o cerca de la parte superior del árbol, este criterio intenta agrupar grandes números de clases similares en algunas características; abajo, intenta aislar clases individuales. La principal desventaja de este criterio es su ineficiencia.

2.2.6.4. Poda.

El algoritmo de Poda es una característica distintiva de CART.

La idea básica es podar aquellos subárboles que produzcan un beneficio pequeño en exactitud. Es de esperar que un árbol más simple podado provoque mayor exactitud en la predicción de casos no vistos ante la presencia de ruido.

Para hacer esto se usa un algoritmo de mínimo coste-complejidad de poda. Se define como:

- Complejidad de un subárbol, T definida como el número de sus nodos terminales (hojas). Se llamará T_c .
- Coste de la clasificación errónea del subárbol T.
- Medida de la complejidad-coste. R_n :

$$R_n(T) = R(T) + \alpha T_c(T)$$

α será el coste de la complejidad por nodo terminal.

La medida coste/complejidad es una combinación lineal del coste del árbol y su complejidad apropiadamente pesada, o sea, es una combinación de exactitud y simplicidad. Al principio si el árbol de decisión es muy grande se suele podar más que cuando el árbol está en niveles bajos. El principal problema de CART es que su cálculo es más intenso.

3. Razonamiento No Monótono y marcos.

La capacidad de razonamiento es una de las cualidades que distinguen a los hombres. Se podría decir que razonar es la capacidad de derivar inferencias de todo tipo (deductivas, inductivas, abductivas, etc.) a partir de un conocimiento dado con el propósito de alcanzar una meta, o resolver un problema. En este contexto, se entiende por inferencias a los procesos (lógicos) mediante los cuales se obtiene una conclusión (un nuevo hecho, proposición, o consecuencia) a partir de unas premisas (proposiciones). Generalmente estos procesos consisten en la aplicación de ciertas reglas. De hecho, muchas veces el hombre puede efectuar inferencias válidas y correctas a partir de conocimiento incierto, impreciso, incompleto, contradictorio o inconsistente.

La lógica nos permite modelar procesos del razonamiento. Pero, como todo modelo, no es más que una aproximación a la realidad que se concreta en determinados aspectos relevantes de esa realidad y que no contempla infinidad de matices y detalles. La lógica clásica (lógica de proposiciones y lógica de predicados) nos da un marco formal para estudiar algunos (aspectos de diversos) tipos de razonamiento. Es por ello que surgen “otras lógicas” que abordan algunos de los aspectos no modelados en la lógica clásica.

En el lenguaje usual o cotidiano se usan mucho “generalizaciones” del tipo, “(todos) los perros y gatos se llevan mal”, “todos los hombres son inteligentes”, “todos los políticos son corruptos”, etc., pero no todo el conocimiento del mundo puede expresarse en forma de afirmaciones universales factibles, o como conjuntos de condiciones necesarias y suficientes para caracterizar un tipo de objetos o personas. Muchas veces hay excepciones. Por ejemplo, si se usa el razonamiento anterior y se tiene:

Todas las aves vuelan

Piolín es un ave

Luego Piolín vuela

en el momento en que se introdujera una afirmación tal como “Piolín es un aveSTRUZ”, “Piolín es un pingüino” o “Piolín está muerto” se estaría obligado a retirar la conclusión de que “Piolín vuela” ya que son excepciones a la afirmación universal “Todas las aves vuelan”.

También se estaría obligado a modificar el conocimiento en situaciones donde el paso del tiempo puede cambiar las características del entorno dando nueva información, que puede ser contradictoria (o no) a la que se tenía anteriormente.

Hay muchos casos prácticos que reflejan razonamiento no monótono, por ejemplo:

La *complección* de bases de datos mediante información negativa implícita. Por ejemplo, en una base de datos sobre líneas aéreas, la forma de razonar sobre la información es asumir que si f es un vuelo y C1 y C2 son dos ciudades, salvo información contraria, asumir \neg conexión (f, C1, C2).

Razonamiento sobre *taxonomías prototípicas* (o jerarquías de prototipos). Este es un caso muy presente en los modelos cognitivos de sentido común. Por tanto, al trasladar las formas de razonamiento matemático al razonamiento del sentido común, se sustituyen las definiciones universales necesarias y suficientes por definiciones de objetos prototípicos a los que se asocia un conjunto de propiedades que caracterizan la mayoría de los componentes de una clase. Dado que estas propiedades admiten excepciones, pueden producirse propiedades retraíbles si aparecen evidencias que afecten a las excepciones.

Razonamiento temporal. Este es uno de los campos en el que el manejo no monótono de hipótesis es el más adecuado. Es habitual operar en un marco cambiante en donde se dan rasgos que caracterizan un entorno que pueden cesar o aparecer bien por causas internas o externas, no siempre conocidas, que obligan a realizar hipótesis sobre el entorno. El razonamiento predictivo debe hacer hipótesis sobre la forma de evolución de los eventos que moldean o definen las características del entorno donde se toman decisiones.

Se puede entender el razonamiento no-monótono como un tipo de modelación del razonamiento de sentido común (Se puede entender por razonamiento de sentido común al proceso que implica el uso de conocimiento cotidiano, de dominio común, para permitir hacer inferencias, planes, predicciones y finalmente obtener conocimiento adicional) que se basa en la identificación de las contradicciones o inconsistencias (del tipo: A y \neg A) y la reacción ante ellas dentro del propio modelo de razonamiento, de forma que el proceso detecta las contradicciones o inconsistencias y revisa el contenido de su propio conocimiento.

Así, puede haber estrategias para la resolución de los problemas que se plantean, por ejemplo:

Incorporar en el procedimiento de inferencia de la instancia la consideración de la existencia de las excepciones, o

Redefinir el concepto universal a efectos deductivos

Por ejemplo si se tiene el caso de las aves:

R: todas las aves vuelan (aquí se aplicaría 2)

P: Piolín es un ave

luego C: Piolín vuela (en este paso se aplicaría 1)

En ambos casos: 1 y 2 se consideraría las excepciones de modo que en:

Durante el proceso de inferencia se evalúen cada una de las excepciones para ver si realmente puedo afirmar que "Piolín vuela"

La definición del "para todo" ahora se tenga en cuenta cada una de las excepciones (como E1: "Piolín es un aveSTRUZ" o E2: "Piolín es un pingüino" o E3: "Piolín está muerto") al momento de usarlo en un proceso deductivo.

Para ello se requiere un razonamiento *metanivel*, que detecta inconsistencias cuando se trata de una excepción y que retrae las afirmaciones supuestamente culpables de su conjunto de creencias.

Es decir que se tienen dos tipos de razonamiento: uno que razona sobre el conocimiento en sí (que es donde se aplicarían 1 y 2, como se ha visto antes) y otro que razona a un nivel superior, controlando al primero o bien considerando un "paso adelante": de inferir A, ¿Se generan inconsistencias con el conjunto de creencias?, ¿Cómo se trata esa posible inconsistencia: A y \neg A?.

3.1. El razonamiento no-monótono en la I.A.

Los primeros trabajos sobre razonamiento no monótono aparecen a finales de los años 70, algunos autores también lo suelen llamar razonamiento revisable. La implementación de un sistema de razonamiento no monótono se plantea por primera vez en Inteligencia Artificial en 1979, cuando Jonathan Doyle presenta su Sistema de Mantenimiento de la Verdad (TMS), como conciencia asociada al control de un proceso de búsqueda. En este sistema se incorpora en forma procedimental una manera de mantener consistentes un conjunto fijo de afirmaciones sobre un dominio del discurso en base a las afirmaciones que se reciben de un proceso principal de búsqueda.

Sobre este concepto trabajaron posteriormente distintos investigadores, como por ejemplo Mc Allester en 1980 y De Kleer en 1986. Este último planteó una variante más declarativa denominada ATMS (Sistema de Mantenimiento de la Verdad basado en Asunciones). En paralelo con estas propuestas, basadas fundamentalmente en implementaciones específicas (*ad hoc*), se han desarrollado propuestas alternativas en el campo de la programación lógica y las bases de datos deductivas (Lógicas por defecto [Reiter, 80], Negación por fallo [Clark, 78]).

Mc Carthy propuso en 1980 el concepto de *circunscripción* como soporte del razonamiento no monótono. Esta idea plantea la redefinición de un predicado o conjunto de predicados en términos de otros predicados que lo implican, y por lo tanto de menor extensión, teniendo en cuenta el contexto del problema que se trate en cada caso. Esta idea ha sido seguida por diversos autores que han ampliado sus capacidades explicativas ([Mc Carthy, 86], [Lifschitz, 85, 86], [Bossu, 85], [Shoham, 87], [Besnard, 88]). El diseño de sistemas de razonamiento no monótono requiere la incorporación de formas de revisión del razonamiento cuya descripción puede plantearse en dos contextos:

Contexto de *lenguaje proposicional*, basado en atributos con dominios finitos de valores predeterminados que incorporan métodos de evaluación y revisión cuyo comportamiento es no monótono.

Contexto de *lenguaje predicativo*. Se plantean formulaciones en lógica de predicados de primer orden a la que se asocia una semántica flexible de algunos predicados.

El estado del conocimiento actual ofrece diversas opciones en la primera línea de tipos de sistemas, no así en la segunda, en donde se han planteado diversos intentos todavía no suficientemente consolidados.

3.1.1. La Negación por Fallo

En este caso se plantea el problema de responder a una pregunta P , dada una base de conocimiento, las respuestas posibles serían P o $\neg P$, si se trabaja con solo dos valores de verdad {Verdadero, Falso}. Sin embargo habría una tercera respuesta alternativa que sería: No sé, cuando en la base de conocimiento no se tienen todos los datos necesarios como para afirmar o negar. (Generalmente este tipo de respuesta no es aceptable, y no se usa). Pero... ¿cómo se llega a esa respuesta? Se presentan tres enfoques alternativos:

Si la base de conocimiento es sólo una enumeración de instancias ciertas (verdaderas) de cada predicado que representa una relación en el dominio. (Se podría ver como una base de datos relacional o como un programa Prolog en el que sólo hay hechos). Si se plantea una n-upla como pregunta P , para una relación dada hay dos respuestas posibles:

Verdadero, en el caso que la n-upla se encuentre en la base.

Falso, como no consta en la base, se asume que lo no registrado en la misma es falso. (Aquí se usa lo que se denomina Hipótesis de Mundo Cerrado [Bossu, 85]).

De manera que la regla de decisión sería: *si se plantea la pregunta P , se intenta encontrar (deducir) P (buscándolo sólo en los hechos); si esto no está, se afirma $\neg P$.* Cuando no se tienen datos sobre todas las n-uplas, no se sabe si esa n-upla verifica o no la relación, por lo que la otra respuesta alternativa sería *No sé*, pero no se considera este tipo de respuesta.

Si la base de conocimiento consta ahora no sólo de relaciones definidas por enumeración de n-uplas, sino también por cláusulas de Horn que ligan predicados unos con otros, definiendo así nuevas relaciones en forma implicativa a partir de otras ya existentes en la base, o que también están definidas en forma implicativa (se podría ver como un programa Prolog con hechos y reglas); el concepto de mundo cerrado se amplía con la hipótesis adicional de que sólo puede afirmarse aquello que es deducible de la definición conceptual. De manera que la regla de decisión sería: *si se plantea la pregunta P, se intenta deducir P (a partir de los hechos y las reglas); si esto no es posible, se afirma $\neg P$* . (Es así como trabaja Prolog).

Al tratar la negación como fallo usando la hipótesis de mundo cerrado, se tiene un comportamiento evidentemente no monótono, ya que si con $\neg P$ se deducen también en la base de creencias Q, R, S, etc.; en cuanto se introduzcan nuevas premisas que hagan deducible P deberán retraerse Q, R, S, etc.

Dada una base de conocimiento como la del punto (2) se podría usar una regla de decisión que diga: *para responder una pregunta P, usar la siguiente condición necesaria y suficiente: $E_1 \vee E_2 \vee \dots \vee E_n \leftrightarrow P$* . Lo que llevaría a que cuando no se cumple $E_1 \vee E_2 \vee \dots \vee E_n$ no se cumpla P, es decir se respondería $\neg P$. Esta formulación lógica, denominada *complección compleja*, la planteó Clark en 1978 y se basa en la redefinición del predicado a partir de sus condiciones suficientes: *si $E_1 \vee E_2 \vee \dots \vee E_n$ es una condición suficiente (\rightarrow) para P, asumirla también como necesaria*, con lo que resulta $E_1 \vee E_2 \vee \dots \vee E_n \leftrightarrow P$. (Este concepto constituye un antecedente de la circunscripción que propuso Mc Carthy).

3.1.2. Las Jerarquías de Marcos.

El concepto de marco, que fue introducido por Minsky en 1975, constituye una forma estructurada de representación prototípica de conceptos. Para definir un marco se definen una serie de variables atributos (*slots*) que caracterizan un concepto y que toman valores en dominios predefinidos. Un concepto se caracteriza por un determinado conjunto de valores para estos atributos. Por ejemplo, el concepto de enfermedad gripeal se podría caracterizar por el siguiente marco:

MARCO	<i>Enfermedad gripeal</i>
Slots	Valores del dominio
<i>Fiebre</i>	<i>Normal, décimas, media, alta</i>
<i>Escalofríos</i>	<i>sí, no</i>
<i>Congestión nasal</i>	<i>Normal, leve, fuerte</i>
<i>Tratamiento</i>	<i>cama, aspirina, aspirina +vitamina C, antibióticos</i>
<i>baja laboral</i>	<i>no baja, 7 días, 15 días, indefinida</i>

Tabla 84: Ejemplo del Marco Enfermedad Gripeal.

Así, una enfermedad gripeal particular, como por ejemplo la gripe asiática, podría caracterizarse mediante determinados valores para los *slots*:

MARCO	<i>Gripe asiática</i>
Slots	Valores del dominio
<i>Fiebre</i>	<i>alta</i>
<i>Escalofríos</i>	<i>sí</i>
<i>Congestión nasal</i>	<i>fuerte</i>
<i>Tratamiento</i>	<i>antibióticos</i>
<i>Baja laboral</i>	<i>indefinida</i>

Tabla 85: Ejemplo del Marco Gripe asiática.

Los atributos pueden clasificarse según dos funciones al razonar (puede haber atributos que tengan ambas funcionalidades):

Atributos discriminantes que sirven para evaluar si un caso dado se ajusta (o no) a las condiciones que definen una clase.

Atributos inferibles que se deducen por defecto una vez que se ha discriminado un caso como correspondiente a una clase.

Las bases de conocimiento organizadas en marcos se pueden estructurar:

Como colección de marcos, cada uno de ellos correspondientes a un prototipo conceptual.

Como jerarquía de marcos, cada marco es o bien un prototipo, o una clase de ellos. Cada nivel incluirá características adicionales respecto del nivel jerárquico previo (o superior) que son representativas de las subclases correspondientes a un nivel.

El proceso de inferencia se hace a partir de un marco pregunta que es donde se especifican los valores para los slots que caracterizan un caso. Normalmente no se refieren a todos los *slots*, sino a un subconjunto de ellos que se considera que es suficiente para evaluar el caso (o bien son los datos de los que se dispone). El proceso de inferencia tiene por objeto decidir qué marcos encajan con el caso que se estudia. Para ello debe aplicarse un proceso de equiparación o *pattern matching* cuyas características generales son las siguientes:

Se evalúan las diferencias entre los valores asumidos en los *slots* de los marcos prototipos y los valores del caso pregunta, determinando en qué grado o medida se ajusta cada par de ellos. Para ello se contemplan diversas opciones:

Puede decidirse en términos nítidos: es o no uno de los valores del dominio.

Puede decidirse en términos borrosos: pueden producirse grados de verificación parcial. Aquí se pueden aplicar técnicas borrosas del tipo de la identificación de la distribución de la probabilidad, o de la acumulación de evidencia, etc. pero al servicio de un modelo de razonamiento no monótono.

Una vez conocido el grado de ajuste de cada *slot*, hay un proceso de síntesis que decide en qué medida se ajusta el marco (visto como una totalidad). Este proceso está abierto a las características de cada aplicación.

Independientemente de las técnicas que se usen en el proceso de equiparación, lo que se decide es si el marco pregunta puede considerarse que encaja en uno o varios prototipos conceptuales de los existentes en la base de datos. Del mismo modo, el proceso de inferencia tendrá sus características particulares dependiendo si la base de conocimiento está organizada como una colección o como una jerarquía de marcos.

Es importante remarcar que cuando se trabaja con inferencia en marcos, una vez que se decidió que el marco pregunta coincide suficientemente con el marco prototipo, el modelo de razonamiento basado en marcos permite asumir por defecto, para el marco pregunta, los valores de los *slots* no definidos en él. El valor de cada atributo asociado a un *slot* puede presentarse explícitamente, mediante uno de los valores del subdominio correspondiente, o bien puede computarse en el momento de la inferencia mediante un procedimiento asociado al *slot*. Así, el valor que se asumirá para completar cada uno de los valores de *slot* del marco pregunta puede obtenerse de una de las dos formas mencionadas anteriormente. Por lo tanto, los marcos son un mecanismo de asunción de hipótesis basadas en el encaje parcial de algunas de ellas en un prototipo predefinido. En particular, el modelo de razonamiento en jerarquías de marcos puede tener comportamiento no monótono cuando en la jerarquía de clases se incluyen valores excepcionales de algunos *slots* ya que en él se puede dar la circunstancia de retracción de hipótesis en la siguiente forma:

Sea un marco a un nivel de la jerarquía cuyas características (*slots*) son A, B, C y D. Es posible que en un nivel superior de la jerarquía se introduzcan dos valores de A y de B que permitan asumir que encajan en el prototipo a ese nivel. Por ejemplo A=a y B=b, el proceso de *pattern matching* se satisface y, por lo tanto, puede inferirse que C=c y D=d. Sin embargo, es posible que un descendiente en esa jerarquía incluya otro concepto discriminante E de modo que se satisfaga para el mismo marco pregunta de manera que en ese nivel de la jerarquía aparece el concepto adicional E=e, pero para el cual los valores de C y D toman valores con alguna excepción respecto de la regla representada por el mismo *slot* en su antecedente jerárquico, por ejemplo C=k. En este caso, el mecanismo de manejo de excepciones en la jerarquía de marcos produce el típico efecto de no monotonía, ya que inicialmente es válida la deducción:

$$A=a, B=b \Rightarrow C=c$$

y sin embargo, al añadir la premisa E=e, es válida otra conclusión:

$$A=a, B=b, E=e \Rightarrow C=k$$

En conclusión, utilizando el modelo de jerarquía de marcos se puede definir un razonamiento basado en excepciones como una forma muy operativa de razonamiento no monótono.

4. Razonamiento aproximado en IA.

Zadeh introdujo la teoría del razonamiento aproximado y otros muchos autores han hecho contribuciones importantes a este campo. Aunque superficialmente pueda parecer que la teoría del razonamiento aproximado y la lógica clásica se diferencian enormemente, la lógica clásica puede ser vista como un caso especial de la primera, aspecto que resulta más evidente cuando se caracteriza a la lógica clásica desde un punto de vista modelo teórico o de semántica formal. En ambos sistemas, se pueden ver a las premisas como inductoras de subconjuntos de mundos posibles que las satisfacen, aunque en el caso de la teoría del razonamiento aproximado esos conjuntos serán subconjuntos borrosos. La inferencia en ambos sistemas está basada en una regla de inclusión: una hipótesis se infiere de una colección de premisas si el subconjunto de mundos posibles que satisfacen la conjunción de las premisas está contenido en el subconjunto de mundos posibles que satisfacen la hipótesis.

La contribución fundamental del razonamiento aproximado es el uso que hace de las variables y la representación de las proposiciones en términos de valores de verdad lingüísticos -subconjuntos borrosos- como valores de esas variables. La lógica clásica sólo usa de modo implícito de idea de variable, en el sentido de valor de verdad asociado a una proposición. Sin embargo, su naturaleza binaria le permite ocultar este hecho, ya que nos podemos referir a una proposición que es verdadera por su denominación, p , y a una que es falsa simplemente por su negación, $\neg p$, evitando así la introducción de una variable V_p cuyo valor sea la valoración de la proposición p . El uso del concepto de variable en la teoría del razonamiento aproximado conduce a tratar dominios que no están dentro del ámbito de la lógica clásica, como es el caso de los problemas que tratan los S.E. borrosos o los controladores borrosos.

La teoría del razonamiento aproximado permite representar también cuantificadores lingüísticos situados entre el *para todo* y *existe* al menos uno clásicos. Esto facilita representar enunciados como la mayoría de los estudiantes eligen asignaturas fáciles, que clásicamente eran tratados de forma reduccionista. Zadeh indicó que un cuantificador como *la mayoría* puede ser representado como un subconjunto borroso sobre un universo de discurso. Los cuantificadores aproximados se usan para representar conocimiento de sentido común.

Una extensión interesante de la teoría del razonamiento aproximado es la posibilidad de tratar con ella conocimiento prototípico. Reiter sugirió una aproximación a la representación de conocimiento de sentido común usando reglas por defecto y Yager lo estudió en el marco de la teoría del razonamiento aproximado. De acuerdo con Reiter, una regla por defecto tal como típicamente los pájaros vuelan, puede ser interpretada así: si un objeto es un pájaro y nuestro conocimiento disponible no es incompatible con que el objeto vuela, entonces asumimos que el pájaro vuela.

Muchos paradigmas de I.A. se basan en el uso de la lógica bivaluada, y la lógica posibilística de Dubois y Prade es uno de ellos. Esta lógica retiene la naturaleza binaria de la lógica clásica, pero permite la asociación de un peso a una proposición, que indica con qué fuerza creemos en la proposición; la medida de la certidumbre que tenemos en ella.

Consideremos el enunciado *El primer niño de María es un chico*. En la lógica posibilística se le asocia a esta proposición un peso $\alpha \in [0,1]$ que indica el grado de creencia en esta aseveración. Debería enfatizarse que asignando este peso a una proposición no estamos diciendo que sea parcialmente verdadera -claramente la proposición de arriba es verdadera o falsa-, sino cuánto creemos en esa premisa. La asociación de pesos a enunciados está presente en la lógica probabilística de Nilsson, que asocia probabilidades a proposiciones, pero la lógica posibilística se diferencia de ésta en que sólo se requiere de los pesos que estén en una escala ordinal. Una ventaja de esta lógica es que permite tratar con reglas en conflicto, permitiendo ordenar las inferencias con respecto a su certidumbre.

La lógica binaria puede ser vista como un caso especial de la teoría del razonamiento aproximado en el cual los conjuntos base tienen dos elementos {T, F} y los grados de pertenencia se restringen a 1 ó 0. La lógica posiblística puede ser vista como una extensión de ésta, en tanto que, aunque se restringen los conjuntos base de valores a dos, T y F, se permiten que los grados de pertenencia sean números en el intervalo unidad.

La LB extiende la lógica binaria permitiendo su formalización en términos de la teoría del razonamiento aproximado. Así, p es *verdadero* alcanzaría la representación V_p es $\{1/T, 0/F\}$, p es falso, V_p es $\{0/T, 1/F\}$ y V_p es $\{1/T, 1/F\}$ indica que el valor de verdad de la proposición es desconocido. En cualquiera de los casos, el conjunto base asociado a la variable el valor de verdad de la proposición p es {T, F}.

5. Sistemas de Información Geográfica.

5.1. ¿Qué es un SIG?

Un Sistema de Información Geográfica es una herramienta informática para el análisis y la visualización de datos relacionados con la geografía. La tecnología SIG integra las operaciones comunes con bases de datos como las consultas y análisis estadísticos con las posibilidades de visualización y análisis geográficos que ofrecen los mapas. Estas posibilidades distinguen a un SIG de otros sistemas de información.

Los mayores retos a los que la humanidad se enfrenta hoy en día, superpoblación, polución, deforestación, desastres naturales, tienen una dimensión geográfica crítica. Dónde crear un nuevo negocio, buscar el mejor suelo para plantar trigo, o mostrar la ruta óptima de un vehículo de emergencia, estos problemas también tiene un componente geográfico. Un SIG permite crear mapas, integrar información, visualizar escenarios, resolver complejos problemas y encontrar soluciones de un modo nunca visto.

5.2. Componentes de un SIG

Todo SIG integra cinco componentes claves: hardware, software, datos, personas, y métodos.

Hardware.

Hardware es el ordenador en donde el SIG opera. Hoy en día, el software SIG se ejecuta sobre una gran variedad de plataformas, desde servidores centralizados a ordenadores personales utilizados en configuraciones aisladas o en red.

Software.

El software SIG aporta las funciones y herramientas necesarias para almacenar, analizar y visualizar información geográfica. Los componentes claves son: Herramientas para la captura y manipulación de información geográfica. Una Base de Datos. Herramientas que soporten las consultas, análisis y visualización de datos geográficos y un interfaz de usuario amistoso.

Datos.

Es posiblemente uno de los componentes más importante de un SIG. Un SIG integrará datos espaciales con otro tipo de datos, pudiéndose incluso utilizar Bases de Datos estándares, empleadas por la mayoría de organizaciones para mantener sus sistemas de información, para almacenar y gestionar la información espacial.

Personas.

La tecnología SIG no aportaría nada sin la contribución de las personas que gestionan el sistema y desarrollan aplicaciones para solucionar problemas reales. Dentro del un SIG se encuentran los especialistas técnicos que diseñan y gestionan el sistema y los usuarios finales que utilizan estos sistemas para su trabajo diario.



Figura 84: Estructura de un SIG.

5.3. Funcionamiento.

Un SIG almacena información sobre el mundo como una colección de capas temáticas que pueden ser enlazadas mediante la geografía (fig. 85). Este simple pero extremadamente potente y versátil concepto ha sido utilizado para resolver muchos de los problemas del mundo real.

Referencias Geográficas.

La información geográfica contiene bien una referencia geográfica explícita, tal como latitud-longitud, o una referencia implícita tal como una dirección, código postal, nombre de carretera. Un proceso automático llamado geocodificación se utiliza para transformar referencias implícitas a referencias explícitas, las cuales le permitirán localizar elementos para utilizarlos posteriormente.

Modelos vectoriales y ráster.

Los Sistemas de Información Geográfica trabajan fundamentalmente con dos tipos de modelos de datos, vector y *ráster*. En el modelo vectorial, la información sobre puntos, polígonos y líneas es codificada y almacenada como una colección de coordenadas x, y. El modelo vectorial es extremadamente útil para describir elementos discretos, pero menos útil cuando se trata de describir elementos que varían continuamente. El modelo *ráster* cubre estas necesidades. Una imagen *ráster* comprende una colección de celdas. Ambos modelos tienen sus ventajas e inconvenientes. Los Sistemas de Información Geográfica modernos son capaces de trabajar con ambos modelos.

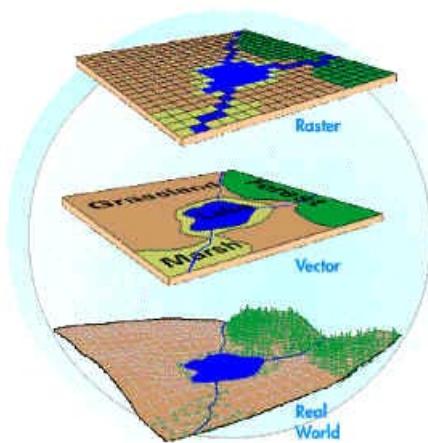


Figura 85: Capas *ráster* y vectorial de un SIG.

5.4. Tareas de un SIG.

Los Sistemas de Información Geográfica de propósito general realizan esencialmente las siguientes tareas o procesos: Carga de datos, manipulación, administración, consulta y análisis y visualización.

Carga de datos.

Antes de que los datos geográficos puedan ser utilizados en un SIG, estos deben ser convertidos a un formato digital adecuado. Al proceso de convertir datos provenientes de mapas en papel a formato digital se le denomina digitalización. Hoy en día muchos de estos datos digitales existen en formatos compatibles con los Sistemas de Información Geográfica.

Manipulación.

Es probable que los tipos de datos necesarios para un proyecto SIG particular necesiten ser transformados o manipulados para hacerlos compatibles con su sistema. La tecnología SIG ofrece muchas herramientas para la manipulación de datos espaciales y eliminación de datos innecesarios.

Administración.

Para pequeños proyectos SIG, puede ser suficiente almacenar la información geográfica en ficheros. Pero cuando el volumen de datos es grande y el número de usuarios utilizando el sistema empieza a aumentar, es mejor utilizar un Sistema Gestor de Bases de Datos para ayudar a almacenar, organizar y gestionar los datos. Existen en el mercado diferentes tipos de Sistemas Gestores de Bases de Datos, pero para un SIG son más útiles los sistemas relacionales. En estos sistemas, los datos son almacenados conceptualmente como una colección de tablas. Los campos comunes dentro de tablas diferentes se emplean para enlazar dichas tablas.

Consulta y análisis.

Una vez que se tiene un SIG funcionando que contiene la información geográfica, se puede empezar realizando una serie de preguntas simples tales como ¿Quién es el propietario de la parcela de la esquina?, ¿Qué distancia hay entre dos lugares?, ¿Dónde se encuentra el suelo industrial dentro de una población?. Y cuestiones analíticas tales como: ¿Dónde se encuentran los lugares en los cuales se pueden construir nuevas viviendas?, ¿Cuál es el tipo de suelo dominante en un robledal?, Si se construye una nueva autopista aquí, ¿cómo afectará al tráfico?.

Un SIG aporta capacidades de consulta simples, basta con pulsar un botón para obtener los resultados, y sofisticadas herramientas de análisis. Los Sistemas de Información Geográfica modernos tienen muchas herramientas de análisis pero hay dos especialmente importantes:

a. Análisis de proximidad: ¿Cuántas viviendas están a una distancia de 1000 m de la autopista?, ¿Cuál es el número total de clientes dentro de un radio de 10 Km. de esta tienda?, ¿Qué proporción de la cosecha de trigo está dentro de la distancia de 500 m. al manantial?. Para contestar a estas preguntas, la tecnología SIG utiliza un proceso llamado “buffering” para calcular las relaciones de proximidad entre distintos elementos.

b. Análisis de superposición: La integración de diferentes capas de datos es realizada mediante el proceso de superposición. Esta superposición, o unión espacial, puede integrar datos de suelos, pendientes y vegetación o propietarios de parcelas con información sobre sus impuestos.

Visualización.

Para muchas de las operaciones de un SIG, el resultado final es visualizado mejor como un mapa o un gráfico. Mientras que los cartógrafos han creado mapas durante toda la vida, la tecnología SIG aporta nuevas herramientas para ampliar la ciencia de la cartografía. La visualización de mapas puede ser integrada con informes, vistas tridimensionales, fotos y elementos multimedia.

Apéndice 2

Incendios Forestales.

En las ultimas décadas, la degradación de la cubierta vegetal en numerosos puntos de la Tierra, y en particular en zonas semiáridas como la cuenca mediterránea, ha alcanzado importantes niveles de gravedad [Ruiz, 97]. Una de las causas principales de ésta degradación ha sido los incendios forestales. De acuerdo con la Dirección General de Protección Civil [AAVV, 98a], *Incendio forestal* es todo aquel fuego que se extiende sin control sobre terreno forestal, afectando a vegetación que no estaba destinada a arder. Se define como terreno forestal (misma fuente), aquel terreno en el que vegetan especies arbóreas, arbustivas, de matorral o herbáceas, sea espontáneamente o procedan de siembra o plantación, siempre que no sean características del cultivo agrícola.

Durante el periodo 1981-1990 [FAO, 92], en el mundo han ardido una media anual de 6.732.000 Has. de bosque, lo que constituye el 0.24% del total, de las cuales, una media de 437.650 Has. corresponden a la vertiente norte de la cuenca mediterránea. España, por sus especiales características geográficas, medioambientales y demográficas, es uno de los ejemplos más lamentables de esta degradación y, baste citar el año 1994, que en este sentido constituyó uno de los más negros de nuestra historia reciente, en el cual ardieron 437.635 Has. [Vélez, 98] en 7.567 incendios [AAVV, 95b], 92 de los cuales fueron grandes incendios, es decir, afectaron a una superficie superior a 500 Has.

Sin embargo, los incendios siempre han sido frecuentes en nuestro planeta, tanto en el espacio como en el tiempo. Los incendios forestales de carácter natural se han producido a lo largo de la evolución de los vegetales desde hace 350 o 400 millones de años. Las primeras evidencias del uso del fuego por los primeros homínidos se remontan a hace un millón o millón y medio de años [González-Alonso, 98].



Foto 1: Incendio en la región de Krasnojarsk (Siberia), aproximadamente 40 km. de frente, 8 de julio de 1995. Fuente: José A. Olivas.

En ecología, se acepta que el fuego ha sido un elemento evolutivo modelador de muchos ecosistemas, y que ha determinado la aparición de numerosas respuestas adaptativas en las plantas pirófitas [De las Heras, 91]. También se ha resaltado el papel fundamental jugado por el fuego en la configuración de determinados bosques de América del Norte y Siberia [Cooper, 61], (Foto 1). La sabana africana, la pradera americana, etc., también deben su origen y/o mantenimiento a incendios periódicos. En la región mediterránea, los jarales, la mancha y la garriga muestran una perfecta adaptación al fuego. Montoya [Montoya, 86] indica que, en zonas donde los incendios se presentan con intervalos inferiores a los precisos para la colonización con *fagáceas*, el pinar de *Pinus pinaster* puede representar la forma óptima de vegetación, dando lugar a lo que Montero de Burgos [Montero de Burgos, 87] llama la *Clímax pirogénica*.

El problema fundamental, consiste en que la perdida de los bosques por los incendios implica la pérdida de sus servicios: protección del suelo, previniendo la erosión, regulando el régimen hidrológico, defendiendo los embalses contra el aterramiento, etc. Todo ello, se ve incrementado en regiones como la mediterránea, en donde las lluvias (muchas veces torrenciales), suelen suceder a la estación en que se producen los incendios. Se pierde su función recreativa, deteriorándose la calidad de las zonas aprovechables para el disfrute. Se destruye el hábitat de las especies cinegéticas, etc.. Por otro lado, la combustión producida en los incendios, emite grandes cantidades de CO₂, CH₄, COS, NH₃, CO, hidrocarburos y otras partículas. El CO y el CH₄ persisten en la troposfera y consecuentemente ejercen un importante influencia en el calentamiento global del planeta debido al efecto invernadero [González-Alonso, 97a]. No podemos olvidar las pérdidas económicas directas (en productos primarios como madera, leña, frutos, etc.), que entre 1961 y 1996 [AAVV, 97b] ascendieron a 1.751 millones de euros, siendo además, su recuperación muy cara y costosa y en muchas ocasiones imposible por las modificaciones que se han producido en el suelo. Por último, hay que lamentar la perdida de vidas humanas que tristemente se producen cada campaña. Entre los años 1961-1997 han alcanzado un número de 257 [AAVV 96, 97a y 98a].

Pero a pesar de que los incendios forestales son negativos para la humanidad, los hemos usado desde siempre por muy diferentes motivos [Vélez, 91]:

- *Como antiquísima herramienta agrícola* (y que continúa utilizándose en la actualidad), bien para ganar terreno al monte y preparar su roturación, como para despejar el suelo después de la cosecha y poder labrarlo nuevamente.
- *Quemas de pastos*, de modo que se regenere y el ganado tenga alimento verde y fresco.
- *Como arma de guerra*: de modo que se destruyan todos los beneficios que este proporcione al país “enemigo”.
- *Recreo*: se trata incendios provocados de forma involuntaria por personas que utilizan el monte como elemento de recreo.
- *Incendiarismo*, típico de algunas regiones, pero que se va agravando extraordinariamente en los últimos años.

Como ya se ha citado, España es uno de los países de Europa en el que la superficie forestal quemada cada año es mayor, aunque hay que considerar, que se trata del 2º país de Europa con mayor superficie forestal (tras Suecia) y el 3º en proporción superficie forestal/superficie total (tras Finlandia y Suecia). Por otro lado las condiciones climáticas que imperan en nuestro país favorecen enormemente su inicio y expansión: un periodo húmedo (otoño-invierno y primavera) que favorece el desarrollo de combustible fino, precediendo a periodos de intensa sequía y que coincide con temperaturas altas (verano) y la aparición de vientos secos del interior, hacen de los bosques un polvorín dispuesto a arder. Por otro lado, la vegetación predominante en España arde con facilidad, lo que facilita su inicio y extensión rápida. Esto fue lo ocurrido en el ya comentado año 1994, cuando precedido por una fuerte sequía, a principios de julio, sobrevino una ola de calor en conjunción con vientos del Oeste que azotaron el sector Este del país, provocando un tremendo número de incendios, buena parte de ellos de grandes dimensiones.

Esta coyuntura ha encaminado a muchos técnicos españoles a especializarse en el campo de los incendios forestales, haciendo que actualmente se encuentren entre los de mayor prestigio a nivel mundial. Así, en el X Congreso Forestal Mundial celebrado en París en 1991, un español se encontraba entre los tres oradores sobre el tema incendios forestales y, en la Segunda Conferencia

Mundial sobre Incendios Forestales de Vancouver de 1997, la situación en la Cuenca Mediterránea, fue presentada también por un español [Vélez, 98].

En un trabajo de estas características, no se puede por menos que hacer un repaso general a la estadísticas de incendios en España, se puede observar un incremento en el número de incendios desde los 64.452 incendios ocurridos entre 1975 y 1984 hasta 132.233 entre 1985 y 1994, es decir, un 205%. De forma paralela, la superficie quemada creció desde 2.120.591 Has. (1975-84) a 2.533.482 Has. (1985-94) lo que supone un incremento del 19.47% [Vélez, 95]. A primera vista, se observa que el incremento en el número de incendios es espectacular. Esto es debido a la tendencia en nº de conatos (incendios cuya superficie afectada ha sido menor a 1000 m²), que ha crecido de forma vertiginosa. Por otro lado y en referencia al resto de incendios, su tendencia es a aumentar su número y que su superficie sea mayor cada año [Moreno, 98]. Aún así, se debe advertir que la siniestralidad de los incendios es sumamente heterogénea y en décadas como la de los 90, ha presentado tanto máximos históricos (1994 ha sido el peor año: se quemaron 250.433 Has. de superficie arbolada, la cifra más alta desde que se creara la base de datos en 1968), como mínimos, también históricos, en los años 1993, 96 y 97 [Vélez, 98] que presentaron cifras comparables a las de hace 30 años.

Tan importante como conocer estas cifras, es conocer su origen, ya que una buena política de prevención de incendios, comienza por prevenir sus causas. Numerosos son los estudios que al respecto se han realizado [Vélez, 98], [Moreno, 98], etc., pero todos coinciden en que la mano del hombre está tras la mayor parte de ellos. Según Moreno [Moreno, 98], la distribución de los incendios por causas para la serie histórica 1974-94, es la siguiente:

- El 2.5% ha sido causado por quemas de pastos, afectando a 2.2% de la superficie total quemada.
- El 3.3% se ha producido por rayos, afectando al 7.5% de la superficie total, el 14.2% a negligencias, suponiendo el 16.7% de la superficie del periodo
- El 37.5% tiene causa desconocida, afectando al 34.1% de la superficie y por último,
- El 42.6% con el 39.6% de la superficie total han sido incendios provocados.

A la vista de esta problemática, el Senado [AAVV, 93a y 95a] emitió dos informes para diagnosticar el problema y proponer asimismo acciones destinadas a hacerle frente. Los incendios forestales, dejan de considerarse como accidentes casuales y sin trascendencia, para merecer la atención tanto de políticos como de técnicos. De hecho, la administración española, junto a la comunidades autónomas, han realizado importantes esfuerzos económicos en investigación y desarrollo en materia de incendios forestales. Vélez [Vélez, 98], habla de los siguientes campos de trabajo:

Prevención: Comportamiento del fuego; Inflamabilidad y combustibilidad; Selvicultura preventiva; Historia del fuego; Análisis sociológico de las causas del fuego; Uso de la teledetección para la predicción del peligro de incendio.

Extinción: Nuevas tecnologías para la detección y comunicaciones; Equipos de seguridad y herramientas; Retardantes y espumas; Equipos aéreos y de tierra; Técnicas para entrenamiento del personal; Sistemas de simulación del fuego y métodos de ataque.

Restauración tras el fuego: Efectos del fuego; Uso de teledetección para evaluar los impactos del fuego; Regeneración natural de especies leñosas tras el fuego; Resistencia de las especies leñosas a diferentes intensidades de fuego; Diseño de reforestaciones y elección de especies para mejorar la resistencia al fuego.

Por último, es importante hacer un repaso al marco legal en el que se encuadran los incendios forestales. Por orden cronológico, aparecen las siguientes normas [AAVV, 98b]:

- Ley 81/1968 de 5 de diciembre, de incendios forestales.
- Decreto 3769/1972, de 23 de diciembre, por el que se aprueba el Reglamento de la Ley de Incendios Forestales.
- Ley 2/1985, de 21 de enero, sobre protección civil.

- Ley 7/1985, de 2 de abril, Reguladora de las Bases del Régimen local.
- Real Decreto 875/1988, de 29 de julio, por el que se regula la compensación de gastos derivados de la extinción de incendios forestales.
- Real Decreto 407/1992, de 24 de abril, por le que se aprueba la Norma Básica de Protección Civil.

Deben también ser considerados los Estatutos de Autonomía de las Comunidades Autónomas, de los que se derivan los Reales Decretos sobre el traspaso de funciones y servicios del Estado en materia de conservación de la naturaleza a las Comunidades Autónomas. Para finalizar, en la Comunidad Europea el reglamento 2158/92 es el que actualmente está en vigor, habiendo sido renovado en 1997, hasta el año 2001.

1. Las nuevas tecnologías.

Las nuevas tecnologías son un conjunto de herramientas que el progreso tecnológico pone a nuestro alcance, y que nos facilitan y nos permiten mejorar nuestro modo de trabajar, e incluso, nos amplían el abanico de posibles actividades orientadas a que la prevención, detección y extinción de los incendios forestales, así como su posterior restauración, sean mucho más eficaces.

A lo largo de los años, un importante número de avances o nuevas tecnologías que han ido surgiendo en el mundo, se han ido incorporando, en la medida de nuestras posibilidades, al trabajo relacionado con los incendios forestales en España. Entre ellas, destacar algunas que actualmente pueden parecer obvias o arcaicas, pero que en su momento constituyeron una importante evolución en el sector. De este modo se pueden citar [Vélez, 98]:

- En la década de los 60, se adquirieron e implantaron los primeros vehículos autobomba: coches todo terreno con chasis Land Rover, con depósitos de 500 l. de agua, mas herramientas, depósitos plegables y motobomba portátil para hacer tendidos de manguera.
- A finales de esta década, se generalizó el sistema de índice de peligro, que partía del modelo canadiense, corregido para considerar las sequías prolongadas y los vientos terrenales desecantes.
- En 1969 se utilizan las primeras avionetas para extinción de incendios (Boeing Stearman), que cargaban 600 l, y se contrató un avión anfibio (Canadair CL-215), primero de este tipo en llegar a Europa.
- En este mismo año, se utilizó el primer helicóptero con tareas de vigilancia.
- Ya en la década de los 70 y con la expansión de la televisión, se utiliza este medio para emitir campañas de concienciación ciudadana y para alertar sobre el peligro de incendio. También se completa la red de comunicaciones provinciales en FM para detección y movilización de medios de extinción.
- En 1971 llegan los primeros Canadair CL-215 adquiridos por España.
- Se utilizan los datos del Instituto Nacional de Meteorología, para prever la evolución del peligro de incendios.
- En 1981 se monta en la Casa de Campo de Madrid, un Simulador de incendios de la U. S. Forest Service.
- En 1984 se utilizan los primeros helicópteros para transporte de cuadrillas.
- En 1989, se firma un contrato para implantar motorización de turbohélice a los aviones adquiridos, para ganar en potencia y seguridad.
- A finales de la década, comienzan a desarrollarse proyectos de investigación sobre inflamabilidad y combustibilidad en le Centro de Investigaciones de Lourizán (Pontevedra) y en le Laboratorio del Fuego del INIA en Madrid.

- Comienza a aplicarse el programa BEHAVE para la predicción del comportamiento del fuego, desarrollado en Estados Unidos.
- Se aplica también el programa CARDIN para la simulación gráfica en ordenador del comportamiento previsible del fuego, desarrollado por la Escuela de Ingenieros de Montes de Madrid.
- Una empresa española desarrolla los extintores de explosión, esferas de plástico que contienen agua y retardante. Portan un detonador térmico de modo que estallan cuando el fuego las alcanza, apagando un área considerable [Vélez, 96].
- Aparecen los refugios ignífugos (*fire shelter*), especie de tienda de campaña de fibra de vidrio aluminizada, que puede ser utilizada en casos extremos.
- Ya en la década de los 90, se utilizan satélites meteorológicos para apoyar la predicción del peligro de incendio,
- El satélite NOAA, se aplica al seguimiento del estado de actividad fisiológica de la vegetación y de las grandes superficies quemadas.
- Se utilizan cámaras de infrarrojos (terrestres y aéreas), para detección de focos calientes.
- Se aplica la telefonía móvil para conectar partes remotas del territorio.
- El G.P.S. (*Global Positioning System*) es utilizado para localizar los aparatos aéreos y para cuantificar las superficies quemadas, inmediatamente después del incendio.
- La aplicación de los S.I.G. (*Sistemas de Información Geográfica*), facilita enormemente la superposición de capas de información georreferenciadas.

Como se observa en los párrafos anteriores, las nuevas tecnologías pueden aplicarse a la prevención, detección, extinción o a la evaluación postincendio y labores encaminadas a su recuperación. De entre las nuevas tecnologías que actualmente se aplican en el sector en estudio, se puede destacar la Teledetección, técnica que nos permite obtener información a distancia de los objetos situados sobre la superficie terrestre, gracias a la existencia de una interacción energética entre dichos objetos y un sensor [Pinilla, 95]. Este flujo energético es una forma de radiación electromagnética, y permite caracterizar las diversas cubiertas que forman la superficie terrestre en función de su comportamiento en las distintas regiones del espectro. Las aplicaciones de la Teledetección al mundo de los incendios forestales son muy diversas y continúan apareciendo día a día nuevos usos y mejoras a las técnicas y aplicaciones ya conocidos. De entre todas ellas, se pueden destacar algunas:

- Estimación del riesgo de incendios forestal, a partir de imágenes NOAA-AVHRR, parámetro de gran importancia, ya que afecta tanto al riesgo de ignición como a la posterior propagación del incendio. Tiene ventajas importantes como su intensidad de medida cada km^2 , muy lejos de ser alcanzada por las redes meteorológicas actuales. Destacar algunos estudios como: [López, 91], [Aguado, 98], [Casanova, 98], etc.
- Estimación de la humedad del combustible, uno de los elementos básicos para conocer el riesgo de incendio forestal. Son muchos los estudios encaminados a este fin [Eidenshink, 90], [Alonso, 96], [Deshayes, 98], etc.
- Cartografía del combustible, que es elemento imprescindible para conocer el desarrollo del fuego. Es además bastante variable en el tiempo, por lo que los mapas tradicionales, pueden quedar anticuados de forma relativamente rápida. Es por ello que en este campo, la Teledetección, se muestra como una herramienta de gran provecho. Trabajos como el de Roberts [Roberts, 93] ó González-Alonso [González-Alonso, 97], pueden resultar interesantes.
- Detección de incendios activos [Martín-Rico, 97], [Lasaponara, 98], etc. La rápida y eficaz extinción del incendio depende de su pronta detección y conveniente localización.

- Seguimiento de focos activos con imágenes NOAA, trabajo en el que tienen gran experiencia los laboratorios de teledetección de las Universidades de Valladolid [Casanova, 98], y Alcalá de Henares.
- Delimitación y cuantificación de la superficie quemada [Martín, 94], [Silgeström, 95].
- Estimación de los daños provocados por el fuego [Caetano, 94], [White, 96].
- Localización de zonas de máxima necesidad de actuación forestal tras los incendios [Ruiz, 97].

Por otro lado, los Sistemas de Información Geográfica (SIG), han supuesto un notable adelanto en todas aquellas disciplinas que trabajan con gran número de datos asociados al terreno. El manejo y análisis de esta información de forma tradicional es engorrosa y muchas veces complicada. Los SIG se están convirtiendo en herramientas de uso habitual, y de ayuda en la toma de decisiones ya que su gran capacidad y cada vez mayor sencillez de manejo, facilita enormemente las tareas de organización de los datos, visualización, superposición, análisis y extracción de resultados. En el caso de los incendios forestales, las aplicaciones de los SIG son muchas y muy variadas. Se pueden destacar:

- **Prevención:** organización previa de las posibles labores de extinción: inclusión de capas con mapas de usos y aprovechamientos, mapa forestal, modelos de combustibles, vías de comunicación, pistas y accesos, modelos digitales del terreno, red hidrográfica y puntos de agua, núcleos habitados, etc.; optimización del diseño de estructuras de prevención, combinando modelos matemáticos con SIG en su realización [Pedernera, 96].
- **Control y seguimiento:** Diseño de modelos de crecimiento del fuego, modelos probabilísticos de su comportamiento, seguimiento del crecimiento y desarrollo del incendio (en combinación con la teledetección).
- **Evaluación y recuperación:** Una vez localizada y cuantificada la zona afectada por el incendio (a partir de técnicas de teledetección, GPS, etc), mediante diferentes cruces con las distintas capas que han sido introducidas en el SIG (mapa de usos de suelo, catastro, modelo digital del terreno, etc.) pueden realizarse una serie de análisis tan amplia como número de capas que se hayan introducido: Evaluación de daños, propiedad del área quemada, superficie por especie, superficie por orientación, alturas, pendientes, etc. Por otro lado, también pueden obtenerse, las zonas con mayor susceptibilidad a la erosión y con mayor necesidad de actuación forestal [Ruiz, 97].
- Por último, es interesante destacar su aplicación a la realización de seguimientos históricos de la dinámica del fuego y sus efectos en el ecosistema.

Junto a éstos, existen otras técnicas, métodos y modelos que se vienen utilizando conjuntamente en las distintas fases de la lucha contra los incendios forestales, algunos de ellos de indudable importancia como ayuda a la toma de decisiones, y que forman un sistema complejo que día a día mejora para conseguir eficazmente su propósito. De entre estos, algunos se encuentran en fase experimental o en proceso de investigación, mientras que otros son ya usados activamente:

- Modelo matemático para determinar las dimensiones de las infraestructuras de protección en las técnicas de selvicultura preventiva [Silva, 98], que permite obtener las distancias mínimas de seguridad recomendadas en diferentes infraestructuras de protección.
- Sistemas integrados para la detección temprana de los incendios [Ollero, 98], sistema inteligente basado la inclusión de diferentes fuentes de información para automatizar la detección fiable temprana de los incendios.
- Sistemas automáticos o semiautomáticos de planificación para la prevención [Anderson, 91, 98], etc., que basados en el riesgo de incendio y el comportamiento del fuego, diseñan la localización y cantidad de recursos contra incendios forestales.
- Modelos para evaluar estos sistemas de planificación para la prevención [Anderson, 98].
- Aplicaciones informáticas como PROMETHEUS [Efichidis, 98] que ayudan en los procesos de toma de decisiones a los responsables de la extinción, de acuerdo con escenarios

predefinidos, o el FOMFIS [Caballero, 98], para la planificación de actividades y distribución de recursos en las labores de prevención de incendios.

- Cámaras de visión térmica, que situadas en puntos estratégicos, son capaces de detectar focos de 1 m² en radios de 10 Km, desarrolladas por la empresa nacional BAZAN en el marco del proyecto BOSQUE [Rayo, 97].
- Modelos que evalúan el riesgo humano para la planificación de las labores de prevención de incendios, basados en evaluación multicriterio.
- Sistemas telemáticos de soporte para centros operativos de lucha contra incendios, en donde mediante redes de control e información compartida, se asumen y comparten las tareas. Están basados en modelos de decisión compartida por cooperación multipersonal, multiexperta [Wybo, 98], lo que actualmente se denomina CSCW (*Computer Supported Collaborative Work*).
- Programas informáticos destinados al inventario y análisis estadístico de los incendios forestales, herramienta de ayuda para la planificación contra los incendios [Dimitrakopoulos, 98].

La superficie media quemada por incendios forestales contabilizada por año en Europa, América del Norte, China y la Federación Rusa es de aproximadamente 6 millones de hectáreas. Medio millón de estas hectáreas se queman en la Europa del Sur. Durante los pasados seis años, los incendios forestales en los países de la cuenca mediterránea han supuesto una incidencia media de aproximadamente 500 millones de hectáreas quemadas por año. Las pérdidas anuales de esta incidencia se evalúan entre 700 y 900 millones de Euros.

El Parlamento Europeo, consciente del problema, ha instado a sus Estados miembros y a la Comisión Europea para que investiguen formas de reducir estas incidencias. La reacción de las distintas administraciones nacionales y regionales de estos países hacia el problema de los incendios forestales ha originado un importante esfuerzo económico encaminado a potenciar, muy especialmente, los medios de extinción. Esta iniciativa ha sido complementada con campañas de prevención y reforestación y ha llevado, así mismo, una concienciación hacia la necesidad de disponer de nuevas tecnologías que faciliten la prevención, detección y extinción de incendios forestales de una manera eficaz.

En las regiones menos pobladas del globo suelen ser los rayos los causantes de los incendios. Sin embargo, en el área mediterránea, donde se cuenta nuestro país, son actividades humanas, utilitarias o recreativas, las que originan la mayoría de los fuegos. El carácter crónico de este fenómeno revela que hay condiciones estructurales, es decir, permanentes, que están en la raíz del problema, estas son, clima, vegetación y población. El clima mediterráneo posee una estación seca y cálida, a veces con fuertes vientos, en la que la humedad de los combustibles muertos se reduce extraordinariamente, facilitando la ignición por focos caloríficos muy pequeños, así como la propagación posterior. La vegetación, en general adaptada al fuego, produce enormes acumulaciones de combustibles muertos en las que el incendio puede empezar y desarrollarse. La población, cuya densidad en las áreas mediterráneas se incrementa precisamente durante la estación seca, invade las áreas forestales en busca de recreo, utilizando, a veces, el fuego. Las poblaciones rurales, aunque en regresión numérica, conservada práctica antigua de emplear el fuego como instrumento de preparación del terreno, tanto para la agricultura como para la ganadería. Sobre la población, como causante inmediata de los incendios, es posible actuar mediante las ciencias sociales para analizar su comportamiento, y mediante la legislación, para orientarlo con espíritu conservacionista.

Hasta ahora, el papel de las nuevas tecnologías está en las actuaciones relativas a la vegetación para analizar su estado y modificarlo de manera que se dificulte la iniciación y la propagación del fuego. En este concepto de nuevas tecnologías se pueden considerar dos grupos, las relativas a proceso de datos y las relativas a nuevos materiales: Las primeras permiten medir gran variedad de parámetros de modo automatizado, manejar y procesar gran cantidad de datos y transmitir información a larga distancia y a gran velocidad. Las segundas permiten obtener resultados en cuanto a duración y rendimiento notablemente superiores a los que proporcionan los materiales tradicionales.

A continuación se hace un inventario de cómo se aplican hasta el momento las nuevas tecnologías tanto para la prevención como para la extinción de los incendios.

1.1. Prevención: predicción del peligro.

El conocimiento de las condiciones en que puede desarrollarse el incendio es esencial para la preparación de medios de combate y para programar acciones persuasivas y disuasorias dirigidas a la población. Este conocimiento precisa de la obtención de datos, su almacenamiento y su procesado [Vélez, 96].

Para la **obtención** de datos se puede contar con:

- La observación meteorológica con estaciones automáticas, dotadas de sensores calibrables, que pueden almacenar datos en memoria y trasmitirlos en tiempo real por radio, cable telefónico o satélite, además de ofrecerlos directamente en pantalla o impresora (INM, DGCN, CCAA). Se cuenta con estaciones automáticas tanto fijas como móviles [Mérida, 98].
- También se ha instalado una estación detectora de descargas eléctricas, con un radio de acción de 400 kilómetros, para identificar los puntos con riesgo de incendios por rayos. El área cubierta principalmente son las montañas del sistema ibérico (centro-este), que presentan la mayor frecuencia de fuegos provocados por rayos. La red de radares del INM proporciona cobertura nacional para este riesgo.
- Detectores por rayos láser, cuyo principio es que su haz detecta columnas de humo [Martínez, 98].
- La teledetección para conocer la situación en grandes áreas mediante imágenes del satélite, además de predicción meteorológica por ejemplo con METEOSAT. Hay varios proyectos en el campo de la teledetección (más adelante se comentan con detalle los programas FUEGO y LATUV), en los que se estudia la aplicación de imágenes de satélites para determinar diferencias de grados de humedad en la vegetación, que muestren incrementos o disminuciones del peligro. El objetivo principal es ayudar a la predicción de situaciones con alta probabilidad de grandes incendios, ya que en ellas se producen los mayores daños. Otro uso de la teledetección es la evaluación de superficies quemadas, y el seguimiento de la evolución postincendio, tema sobre el cual ya hay una serie de experiencias positivas y que requiere el desarrollo de metodologías adaptadas a las necesidades de la estadística de incendios forestales.
- Cámaras de visión nocturna e Infrarrojos, que como ya se ha indicado, alertan automáticamente al detectar un foco caliente [Mérida, 98]. Se complementan con cámaras de video, de modo que pueda comprobarse de forma remota la veracidad de la alarma. También se cuenta con un sensor de infrarrojos montado sobre una unidad móvil. Como experiencia de la Junta de Andalucía, se está empleando la Red Eléctrica para colocar sensores infrarrojos y cámaras CCD en las líneas de alta tensión, aprovechando la línea de fibra óptica para realizar las comunicaciones.
- Torretas o casetas de vigilancia, el sistema tradicional y muy eficaz [Martínez, 98].
- Detección móvil terrestre con ventajas como que puede realizar un primer ataque al foco o el efecto disuasorio ante posibles pirómanos; o aérea que gracias a los GPS que portan, son capaces de situar con precisión el foco. Cuentan entre sus ventajas el cubrir grandes zonas por unidad de tiempo y que por su situación, puede realizar labores de coordinación [Martínez, 98]. Los aviones OPTICA que contrata la DGCN, cuentan con cámaras de video, que envían imágenes en tiempo real del foco o incendio, a través de telefonía móvil.

Para el **almacenamiento** de datos se puede contar con:

- Las bases de datos, manejables actualmente con PC. La base de datos sobre incendios forestales en España (EGIF) es la más grande de Europa, habiéndose iniciado en 1968, al aprobarse la Ley de Incendios Forestales. Es también la más completa, con el mayor número de datos sobre cada incendio. Tiene carácter descentralizado, de manera que puede manejarse

a niveles nacional, comunitario, provincial, comarcal o municipal según se requiera para los trabajos de planificación. A nivel nacional, los datos grabados en Dbase IV son trasladados a INFORMIX como base de datos relacional.

- Los sistemas de información geográfica, que crean bancos de datos en los que puede aparecer, sobre cartografía digitalizada, la información conveniente acerca del estado selvícola de los montes, especies, modelos de combustibles, etcétera. La base de datos ARC/INFO y el nuevo Inventario Forestal son los instrumentos que la Dirección General de Conservación de la Naturaleza está poniendo marcha con esta finalidad.

El **proceso de la información** para predecir el peligro cuenta con desarrollos como:

- Los modelos de comportamiento del fuego que describen la evolución de incendios de origen puntual a partir de datos de combustibles, meteorología y relieve. En España se adaptó hace diez años el modelo BEHAVE, desarrollado por el *U. S. Forest Fire Laboratory*. El empleo de este sistema requiere la identificación de modelos de combustibles, según una clasificación normalizada, trabajo realizado por el ICONA con la colaboración del INIA. Están ya disponibles las claves correspondientes a los montes de toda España.
- Programas de ordenador, que permiten en utilizar los modelos. Esta disponible el BEHAVE en español para hacer predicciones acerca de la velocidad de avance del fuego, su intensidad calorífica, la altura de las llamas y su forma, así como orientar para tomar decisiones acerca del empleo de medios aéreos como apoyo a los de tierra.

En los últimos años la Universidad Politécnica de Madrid ha desarrollado un programa de simulación, denominado CARDIN, que permite prever la evolución más probable de un incendio. Se completa con DIGICAR, programa para la creación de la bases de datos geográfica, y con GFUEGO, programa para la gestión de los medios de extinción, pretendiendo que el conjunto constituya un Sistema Experto. Se ha preparado para que pueda leer datos del SIG TERRASOFT, disponible en varias comunidades autónomas, con mapas de modelos de combustibles, necesarios junto con los datos topográficos y meteorológicos para hacer la simulación. Se pretende la compatibilidad de CARDIN con ARC/INFO. Se tiene en estudio la utilización de los datos del Inventario Forestal y del Mapa Forestal para generar mapas de modelos de combustibles. Mientras tanto el mapa forestal sobre ARC/INFO es una herramienta de uso ordinario en la central de operaciones de la Dirección General de Conservación de la Naturaleza para la toma de decisiones sobre movimiento de medios aéreos de cobertura nacional.

El sistema PIROMACOS, se apoya en CARDIN y proporciona, junto a la simulación gráfica, el despliegue óptimo de los medios de extinción, con las rutas críticas según las pistas en el monte y, mediante retroalimentación, va evaluando las necesidades sucesivas de nuevos medios. Como parte de sistema, se emplea un módulo de cálculo del índice de peligro a partir de la probabilidad de ignición del combustible fino muerto y teniendo en cuenta la fuerza y dirección del viento, lo que permite definir tres estados de peligro, prealerta, alerta y alarma, con sus correspondientes consecuencias en cuanto a medidas preventivas y disponibilidad de medios de extinción.

1.2. Infraestructura: detección y comunicaciones.

Las medidas preventivas pueden reducir el número de incendios, pero sería utópico pretender que no se produjera ninguno. Por ello, es preciso disponer de medios de ataque y de un sistema de movilización. En este último campo la tecnología actual ofrece numerosos avances, tales como:

- La instalación de centrales de operaciones. En ellas se pueden reunir sistemas complejos de comunicaciones generales, tel., télex, telefax y coordinar redes propias de radio en F. M., para uso local, y en H. F. para uso regional. Estas centrales desempeñan el papel de coordinadores de medios, con el valioso instrumento de los ordenadores personales, mediante los cuales se puede utilizar los programas citados de predicción del comportamiento del fuego, los bancos de datos sobre incendios y los inventarios de medios para gestionarlos con la máxima productividad.

- Las estructuras de vigilancia terrestre, constituidas por las clásicas torres y casetas en puntos dominantes, desde las que detectan los incendios personas conocedoras del terreno. La utilización de estructuras de madera laminada y de otros productos derivados de la madera permite mejorar la habitabilidad de estas instalaciones, reduciendo su impacto en el paisaje. La escasez de personal para labores de detección comienza a hacer interesante el empleo de sensores de infrarrojos combinados con cámaras de vídeo. Con carácter experimental, la D. G. C. N. ha montado dos estaciones con el citado equipamiento en las que se están desarrollando los programas de identificación de alarmas y transmisión a distancia (radio o teléfono con vídeo lento), compatibles con la facilidad e manejo. El radio de cobertura está entre 10 y 15 kilómetros. Asimismo se ha construido una unidad móvil para su desplazamiento a los incendios y seguimiento de puntos calientes en el perímetro. Otras iniciativas en este terreno han sido las de la Junta de Andalucía, con otras dos estaciones y la de Red Eléctrica que está experimentando el empleo de sensores en líneas de alta tensión, enviando la señal a través de fibra óptica, como una colaboración a la protección del entorno en que se sitúan sus instalaciones.
- La vigilancia móvil terrestre, es imprescindible para completar la información dada por los puntos fijos. La tendencia actual es combinarla con el primer ataque. Para ello, la D. G. C. N. ha desarrollado vehículos todo terreno con equipo desmontable de cisterna (300 - 400 litros) y motobomba para utilizarlos como coches patrulla en las áreas de mayor riesgo, añadiendo sus funciones la preventiva de la disuasión al hacerse visible al público.
- La vigilancia aérea permite completar la cobertura de vigilancia en comarcas muy montañosas donde se producen muchas áreas ciegas, no divisables desde ningún punto fijo. Para esta vigilancia el empleo de equipos GPS es imprescindible, al permitir situar exactamente el foco por coordenadas. Estos aparatos también son interesantes para la gestión de los medios y para la medida rápida de superficies quemadas. La detección aérea simple se ha visto mejorada por la posibilidad de usar también la televisión y los infrarrojos. Aeronaves dotadas de cámaras de infrarrojos pueden realizar mapas de incendios en progresión, "viendo" a través de masas de humo y proporcionando valiosa información a los combatientes. Estos equipos pueden, incluso, transmitir esos mapas a las centrales de operaciones vía telecopiadora. Las cámaras de televisión bajo avión o helicóptero pueden, igualmente, transmitir imágenes del desarrollo del fuego, utilizables para organizar su extinción. La D. G. C. N. utiliza desde hace varios años de modo sistemático pequeños aviones (aeronaves espía) con cámara de vídeo que envían imágenes por telefonía móvil a las centrales de operaciones. Esto permite evaluar mejor las situaciones de peligro y tomar decisiones con mayor fundamento que cuando sólo se dispone de datos verbales. Con este sistema se han ahorrado cientos de horas de vuelo, al evitar falsas alarmas o peticiones exageradas de medios. La telefonía móvil permite enviar imágenes entre dos puntos cualesquiera de España. Por microondas se puede enviar imagen en movimiento en un radio reducido, lo que es una gran ayuda para el director de extinción situado en el propio incendio. También se están considerando los aviones no tripulados, aunque su coste los hace poco interesantes por el momento.
- Las unidades móviles de meteorología y transmisiones. Son vehículos todoterreno en los que se montan centrales de comunicaciones con repetidores para las redes de F. M., A. M. y H. F. para facilitar los enlaces en el campo, especialmente en los grandes incendios. Las unidades desarrolladas están equipadas, además, con una estación meteorológica automática y un ordenador para predecir la evolución del fuego. Estas unidades se empiezan a equipar con receptores de imágenes por microondas para recibir las que envíen las aeronaves espía. Las aplicaciones más recientes en el campo de las comunicaciones son las que permiten el seguimiento de móviles sobre un Sistema de Información Geográfica y el envío de mensajes a esos móviles vía satélite. El seguimiento de móviles consiste en enviar la señal GPS a través de radio a un ordenador situado en la central de operaciones, facilitando el control de trabajos y de tiempos en un ámbito provincial. Por vía satélite (INMARSAT) se puede enviar a un avión en vuelo órdenes escritas para corregir falsas alarmas o modificar su objetivo, así como conocer en todo momento su posición y recorrido sobre un Sistema de Información Geográfica. La D. G. C. N. está instalando este sistema en sus aviones anfibios CL-215, dado que pueden ser enviados a cualquier punto del territorio nacional.

1.3. Infraestructura: Equipo terrestre para extinción.

Se coincide en reconocer que el mejor equipo terrestre de extinción es una brigada compuesta por personal adiestrado, en buena forma física y con experiencia en el combate de incendios. Sin embargo, su labor se puede ver facilitada y su eficacia mejorada utilizando los medios disponibles, tales como:

- El equipo personal de protección. Las nuevas fibras no inflamables permiten emplear ropa de trabajo que proteja suficientemente al atravesar zonas de matorral, sin estorbar a los movimientos y, a la vez, evitar las quemaduras cuando caen pavesas sobre la tela. Cascos, gafas, guantes y botas componen con dicha ropa un equipo adecuado para esta actividad. Para las situaciones de gran concentración de humos, se está proporcionando mascarillas al personal. Asimismo, se empieza a considerar la posibilidad introducir el "refugio antifuego" (*forest fire shelter*), especie de tienda aluminizada que se lleva plegada en el cinturón y que ha salvado vidas en Estados Unidos y Australia al ser cercado personal por el fuego.
- Los vehículos todo terreno contra incendios. En España existen varias fábricas de excelentes chasis todoterreno, carrozables para incendios forestales, con cisternas de tres mil a tres mil quinientos litros, mas bombas, mangueras y material auxiliar. Los diseños más recientes poseen un alto poder de penetración en terrenos muy quebrados y permiten transportar agua a donde haga falta, tanto para el combate como para la liquidación del incendio, ya que la red de caminos, por razones económicas o ecológicas, nunca puede ser tan densa como los trabajos de extinción podrían necesitar. Estos vehículos, asimismo, se están adecuando actualmente para lanzamiento de espumas e incluso de retardantes de largo plazo para obtener mayor rendimiento del agua transportada.
- Una nueva forma de aplicación de los retardantes la constituyen los extintores de explosión, creados y desarrollados por empresas españolas. Se trata de esferas de plástico llenas de agua con retardante que se colocan cerca de la línea de fuego, y al explotar dispersan su contenido apagando un área considerable.
- La maquinaria pesada (*bulldozer*) es fundamental para la extinción, tanto para la apertura de cortafuegos, como para el ataque directo. Se están desarrollando equipos que combinan el trabajo de remoción del terreno con el transporte de una cisterna de 2.000 a 3.000 litros para remate del perímetro del incendio. Recientemente han aparecido en el mercado tanques modificados para transportar 10.000 litros de agua y lanzarlos con un cañón. Van dotados con una pala “*dozer*” para ir derribando el matorral y se mueven sobre orugas.
- La normalización y homologación de materiales. La concurrencia de varias administraciones en la lucha contra los incendios, obliga a la normalización de materiales para hacer posible las acciones conjuntas. Asimismo, la compra de estos materiales precisa la elaboración de directrices comunes y de procedimientos de homologación por obtener suministros con la calidad adecuada. La aplicación de sistemas de control de calidad con estos fines es otra de las aportaciones de la ingeniería para mejorar la eficacia de los servicios que combaten los incendios. De estas actividades se ocupa el Comité de Lucha contra Incendios Forestales, que coordina a las administraciones central y autonómicas.

1.4. Infraestructura: Equipo aéreo para extinción.

A los 28 años de emplearse por primera vez en España aviones contra los incendios forestales, es interesante comprobar cómo se ha extendido y diversificado la utilización de medios aéreos. Durante 1995 operaron 187 aeronaves, dando cobertura aérea a todo el territorio nacional. Las tendencias actuales en el empleo de medios aéreos de extinción son las siguientes:

- Los aviones anfibios se han mostrado decisivos para países como España, con numerosas rías y golfos a lo largo de las costas, así como una densa red de embalses en el interior donde pueden cargar agua. Estos aviones constituyen el núcleo central y permanente del ataque aéreo contra incendios. La gran ventaja de estos aviones que su capacidad por operar sin más infraestructura de tierra que la ordinaria proporcionada por los aeropuertos. Ello permite

utilizarlos en cualquier punto y trasladarlos de una zona a otra sin más limitaciones que las de mantenimiento.

Sin embargo, la flota extendida por los países mediterráneos y Canadá se enfrenta con el problema de la remotorización por obsolescencia de los motores de pistón. Por ello, España ha optado en 1989, anticipándose a los demás países, por la modernización de la flota española mediante utilización de motores turbohélice, que incrementan la potencia y velocidad de los aparatos y disminuyen sus necesidades de mantenimiento. Actualmente son 16 los aviones Canadair CL-215T que operan en España (Foto 2).



Foto 2: Canadair CL-215T actuando en un incendio en Galicia.

Fuente: *Lufthansa Magazin* 12 (Diciembre), 1999.

- Los aviones pesados de carga en tierra. Estos aparatos, ampliamente utilizados en Norteamérica, empiezan a estar disponibles en Europa. Para algunas regiones, estos aviones podrían dar un servicio interesante, aunque tienen la contrapartida de precisar instalaciones de carga de agua y retardantes en bases situadas en aeropuertos, lo cual en un país con el riesgo tan extendido como España, supone una limitación muy fuerte. Se han utilizado en los cinco últimos años debido a la fuerte sequía que ha vaciado los embalses, disminuyendo la operatividad de los aviones anfibios. Se han probado el DC-6, el DC-7 y el C-130, siendo éste último que el de mayor capacidad y maniobrabilidad.
- Los aviones ligeros de carga en tierra están dando un servicio muy interesante en todas las áreas forestales en España, ya que sus escasas exigencias de infraestructura permiten instalarlos en bases situadas en el propio monte. La capacidad de los aviones utilizable a ido creciendo y ya existen muchos aparatos disponibles con capacidad entre 1.500 y 2.500 kilos de carga de agua con retardantes.
- Los helicópteros son la aportación más reciente de la aeronáutica a la lucha contra los incendios. Su elevado coste de operación no los hace adecuados para la vigilancia, pero su velocidad y capacidad de tomar tierra en sitios muy diversos los hacen insustituibles para el transporte rápido de cuadrillas y el ataque en lugares de difícil acceso por tierra, así como para las funciones de coordinación. El lanzamiento de agua también es factible con los helicópteros, bien con helibalde (*bamby*), bien con depósito ventral llenado con bomba y manguera. Este sistema fue introducido por ICONA con éxito en España durante la época de peligro de 1988. Los helicópteros procedentes del este (*Sokol, Kamov, Mi-2, Mi-8, Mi-17...*) por su excelente relación prestaciones/coste han permitido generalizar el uso de este medio, también son muy usados los franceses Aeroespaciale HA-16 (Foto 3).



Foto 3: Helicóptero HA-16 en la base de Sta. Cruz, Lugo.

Fuente: José A. Olivas, 1988.

- Los retardantes son una aportación de la industria química que ya se considera imprescindible para los aviones de carga en tierra. La existencia de producción española de retardantes de largo plazo hace que se posea ya una gran experiencia en su aplicación. En 1985 se comenzó a utilizar espumógenos con aviones anfibios y también se emplean con helicópteros y con los aviones de carga en tierra.

1.5. Formación del personal.

Este largo inventario, en el que por supuesto, no se ha entrado en detalles, muestra la variedad de medios que la tecnología actual pone a disposición de los encargados de combatir un problema de la magnitud alcanzada por los incendios forestales. Sin embargo, todos los sistemas y equipos, por muy sofisticados que sean, no valen más que las personas que los manejan. Por ello, la aplicabilidad de las ofertas de las nuevas tecnologías no pueden ser evaluadas simplemente en

función de las prestaciones que pueden dar, sino teniendo en cuenta además la estructura de la organización de la defensa contra incendios forestales y la formación del personal disponible. La formación, garantizada a través de la profesionalización, es un tema básico para que la introducción de nuevos métodos y equipos sea verdaderamente efectiva y rentable.

2. Programa FUEGO.

FUEGO es un programa para desarrollar un sistema espacial dedicado específicamente a la lucha contra los incendios forestales. El Programa FUEGO está siendo llevado a cabo por el Grupo FUEGO, consorcio industrial integrado por: *Ingeniería y Servicios Aeroespaciales, S.A.* (INSA), España. Líder del programa. *Officine Galileo*, Italia. Participa en la definición del sensor. *Sema Group*, Francia. Desarrolla las herramientas de simulación. *Instituto Nacional de Técnica Aeroespacial (INTA)*, España. Participa en las campañas de vuelo de las quemas controladas en campo y en la definición de sistemas auxiliares. Los asesores científicos del Grupo FUEGO son: *Instituto Nacional de Investigaciones Agrarias (INIA)*, España: participa en los ensayos en túnel. *Centro de Investigaciones Forestales de Lourizán (CIF)*, España: participa en las quemas controladas en campo.

El sistema FUEGO consiste en una constelación entre 9 y 12 satélites en órbita baja, equipados con sensores infrarrojos y procesadores de gran capacidad para la DETECCIÓN y MONITORIZACIÓN de incendios forestales y para la EVALUACIÓN de los daños que éstos han causado. El sistema debe permitir la cobertura de todas las áreas de alto riesgo de incendios forestales a nivel mundial y específicamente la Europa Mediterránea, EE.UU., Australia y Sudamérica. El Programa FUEGO se inició oficialmente el 15 de Noviembre de 1996. El objetivo es desarrollar un servicio operativo para el verano del año 2.002.

El Programa FUEGO surge, inicialmente, desde el convencimiento de que la disponibilidad de información en tiempo real sobre los incendios o conatos con el adecuado nivel de proceso requerido, proporcionará una herramienta eficaz que facilite el proceso de decisión de los responsables de su extinción y, por tanto, el control del suceso. La tecnología actual en Europa, sobre todo en materia de procesadores de gran capacidad y de sensores infrarrojos, permite llevar a cabo este sistema de una manera abordable económicamente.

El sistema FUEGO se concibe sólo si se ofrecen productos que cubran las necesidades de detección y monitorización de los usuarios potenciales a precios aceptables. El sistema FUEGO pretende estar al servicio de los responsables de la lucha contra los incendios forestales. Se pretende recoger desde el inicio los requisitos y necesidades de estos usuarios para proponer un sistema útil. Los usuarios son parte activa del programa mediante Comités y Grupos de Usuarios. Con el fin de recabar los requisitos de los usuarios, se ha construido un cuestionario que recoge los requisitos de los usuarios en materia de detección, monitorización y actividades posteriores al incendio.

Las conclusiones de la Primera Conferencia de Usuarios del programa son las siguientes:

- Areas de cobertura: la superficie aproximada a cubrir por el sistema es de 12 Mha. para España, 10 Mha. para Italia, 3 Mha. para Portugal y 7 Mha. para Chile.
- Tiempo medio de detección: Es la clave para el control del incendio. Un tiempo medio de detección de 15 minutos es aceptable para los usuarios.
- Resolución de la línea de fuego: un tamaño de pixel de 30-40 m. es adecuado.
- Distribución de los datos: la información FUEGO debe enviarse a los centros regionales y provinciales.
- Necesidad y precisión de geo-localización: una precisión de 300 a 500 m. es aceptable en detección. Precisiones del orden de 30 - 40 m. son precisas en monitorización.

Es importante la disponibilidad del sistema fuera de las épocas de riesgo, cuando los recursos humanos y materiales son escasos o incluso inexistentes. FUEGO debe ser compatible con la infraestructura de lucha contra los incendios forestales existente.

El factor de riesgo principal para la definición de las áreas de cobertura es el viento. Otros factores importantes son el meteorológico y los humanos, mientras que los combustibles y la topografía son factores de segundo orden. Los mapas de riesgo deben ser flexibles, con capacidad para ser modificados rápidamente según los cambios de los factores más influyentes. Los cambios climáticos deben incluirse en el sistema diariamente (áreas a vigilar). Los cambios estructurales, como los combustibles, deben incluirse bianualmente. El sistema principal de detección se basa en torres fijas con observadores, lo que genera muchos problemas de personal y de visibilidad durante la noche. Los niveles aceptables de falsas alarmas varían entre el 5 y 20%, dependiendo de las regiones. En regiones con gran número de incendios las falsas alarmas representan un aspecto muy crítico. La monitorización debe considerarse asociada con herramientas de simulación y otro software para gestión del fuego y operaciones nocturnas. Las actividades posteriores a los incendios deben centrarse en la evaluación del perímetro de incendio y el análisis del daño ecológico causado.

La estructura de comunicaciones actual se basa en redes de radio (80 y 140 MHz). Esta estructura tiene problemas de saturación, interconectividad entre las diferentes regiones y provincias y la falta de canales para la comunicación con los medios aéreos. La comunicación entre centros se efectúa con la red telefónica nacional, y de ahí el potencial de los nuevos sistemas de comunicaciones por satélite. Los usuarios reconocen la importancia de disponer de SIG (Sistemas de Información Geográfica) para soporte y gestión de las actividades de lucha contra los incendios forestales. Actualmente, estos sistemas no están disponibles en la mayoría de las regiones. La experiencia con nuevas tecnologías (p.e., torres con Infrarrojos o de detección de humos) no ha sido positiva en general, aunque los usuarios manifiestan la necesidad de estar al corriente de los nuevos desarrollos.

El sistema FUEGO está formado, inicialmente, por nueve satélites a 700 km. de altura en tres planos inclinados 47,5°, de modo que se maximice el tiempo de revisita en latitudes próximas a los 40°. La figura 86 resume los principales parámetros de observación del sistema FUEGO.

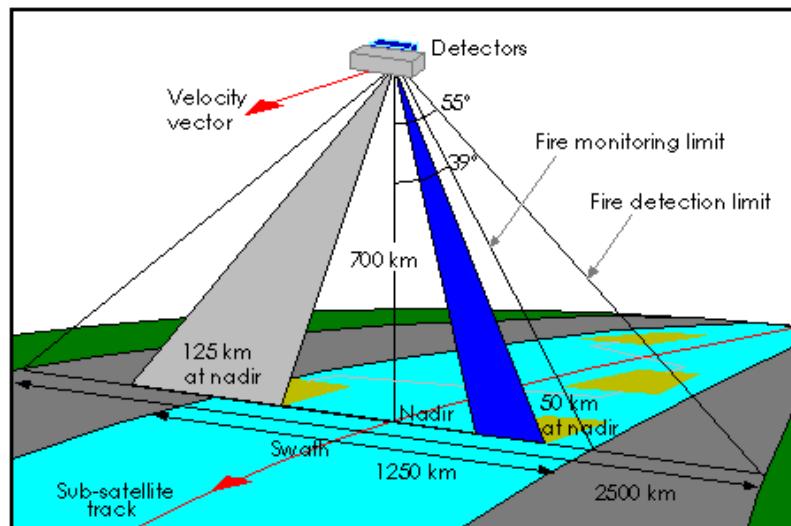


Fig. 86: Principales parámetros del sistema FUEGO.

El sistema y la configuración propuesta proporcionan un tiempo de revisita de 35 minutos, lo que significa unos 18 minutos de tiempo medio de detección. Otra solución que está siendo estudiada para disminuir el tiempo medio de detección es agrupar los satélites de manera que se obtenga un tiempo de revisita más bajo en los períodos y en las horas del día de máximo riesgo de incendios, mientras que en los períodos y horas de menor riesgo el tiempo de revisita sea mayor. Si las órbitas se sitúan en planos con inclinación orbital de 132,5° (órbitas retrógradas), el tiempo medio de revisita que se obtiene con esta configuración es de 19 minutos, es decir, menos de 10 minutos de tiempo medio de detección. Actualmente, la configuración final de la constelación está en estudio.

La carga útil de FUEGO está formada por los siguientes elementos:

- **Detectores infrarrojos.** Los detectores infrarrojos en los rangos de los 3 a 5 mm. y de 8 a 10 mm. se consideran útiles para la detección y monitorización de incendios. Además, otros canales en el visible y el infrarrojo cercano (NIR) pueden proporcionar otra información de interés para el filtrado de falsas alarmas. La necesidad de disponer de una resolución en tierra muy alta en monitorización y la limitación del número de elementos que se pueden concatenar en la misma barra de detectores han obligado a disponer de dos instrumentos distintos abordo: el sensor para detección y el haz estrecho para monitorización.
- **Sensor para detección.** Su principal función es la detección de incendios. Para ello es preciso disponer de un campo de vista amplio (2.500 km.). La resolución de este sensor está dada por el mínimo conato que se desea detectar.
- **Haz estrecho de monitorización.** Su función principal es la monitorización de incendios en curso. Para ello requiere una resolución espacial del orden de los 30 a 60 m., superando posibles problemas de saturación del sensor. La resolución temporal requerida para este haz está en torno a los 90 minutos. Otra función de este haz es completar la observación las áreas de vigilancia (AOS, *Areas of Surveillance*) definidas por los usuarios, cuando el haz de detección no puede cubrirlas.
- **Espejos.** La configuración en dos haces requiere la existencia de dos espejos con ejes de rotación paralelos e independientes. Estos espejos permiten observar las AOS previamente identificadas por los usuarios.
- **Procesador.** La principal tarea del procesador de la carga útil FUEGO es la compresión de los datos adquiridos de modo que se obtengan unas velocidades de transmisión compatibles con las bandas de comunicaciones seleccionadas. Por ello, el procesador consistirá básicamente en un compresor dedicado.

El sistema debe cubrir todas las áreas de alto riesgo a nivel mundial entre los paralelos 50° N y 50° S. Se han definido tres niveles de servicio como se muestra en la figura 87.

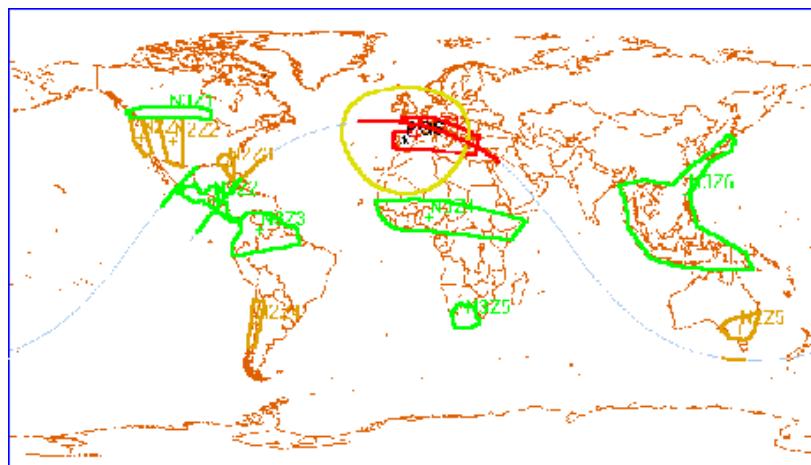


Figura 87: Cobertura del sistema FUEGO

El nivel 1 (rojo) incluye los países mediterráneos europeos (España, Italia, Portugal, Francia y Grecia). El nivel 2 (marrón) incluye los Estados Unidos (California, Florida, ...), Sudeste de Australia y Chile. El nivel 3 (verde) incluye otros países en el mundo con graves problemas de incendios forestales (no se han dibujado todos). Los parámetros orbitales están optimizados para obtener la máxima resolución temporal en los niveles 1 y 2. El nivel 3 tendrá unos valores de resolución temporal algo menores. El segmento terreno de FUEGO se está diseñando basado en una única estación de tierra principal (PGS, *Primary Ground Station*) a cargo de los datos de telemetría y del control de la constelación y de múltiples estaciones de usuario primarias (PUS, *Primary User Stations*) localizadas a nivel regional y provincial y operadas directamente por los usuarios del sistema. La necesidad de una estación de tierra secundaria está siendo estudiada. La infraestructura terrena se completa con terminales de Usuario (FUT, *FUEGO User Terminals*), que son pequeños ordenadores que permiten la recepción de los datos desde los puestos de mando

avanzados. FUEGO debe producir dos productos distintos: uno para detección y otro para monitorización. **Producto de detección**. Un ejemplo del producto FUEGO para detección se muestra en la figura 88.

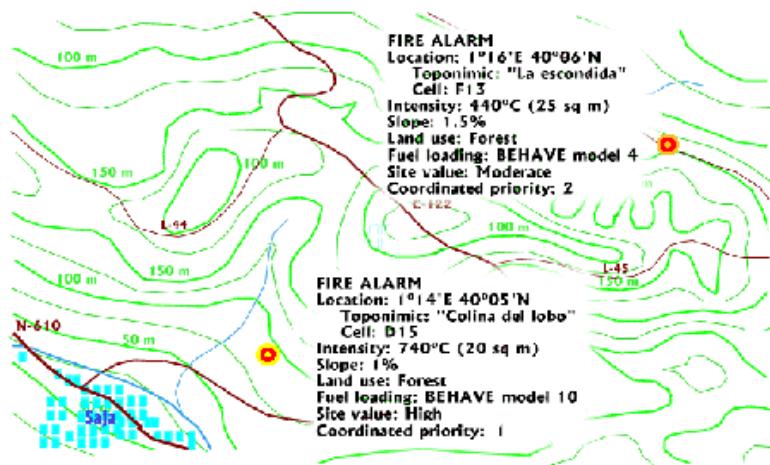


Fig. 88: Ejemplo de producto FUEGO de detección

Cuando el sistema detecta un incendio aparece una alarma sonora en las instalaciones del usuario. La pantalla del terminal FUEGO (FUT) muestra gráficamente la localización de la alarma(s), proporcionando las coordenadas en distintos sistemas de referencia y las dimensiones y temperatura del conato. El sistema de información geográfica (SIG) residente en el FUT añade otra información relativa al terreno tal como la pendiente, la carga de combustible del terreno, el valor asignado al lugar u otros. **Producto de monitorización**. El producto de monitorización se proporciona al usuario automáticamente para incendios superiores a 25 Ha. y bajo demanda para incendios inferiores a esa cifra. El producto de monitorización muestra la temperatura del perímetro de incendio codificada en color y superpuesto con información SIG que proporcione mapas de accesos, líneas de nivel, carga de combustibles, etc, tal como se muestra en el siguiente ejemplo (figura 89).

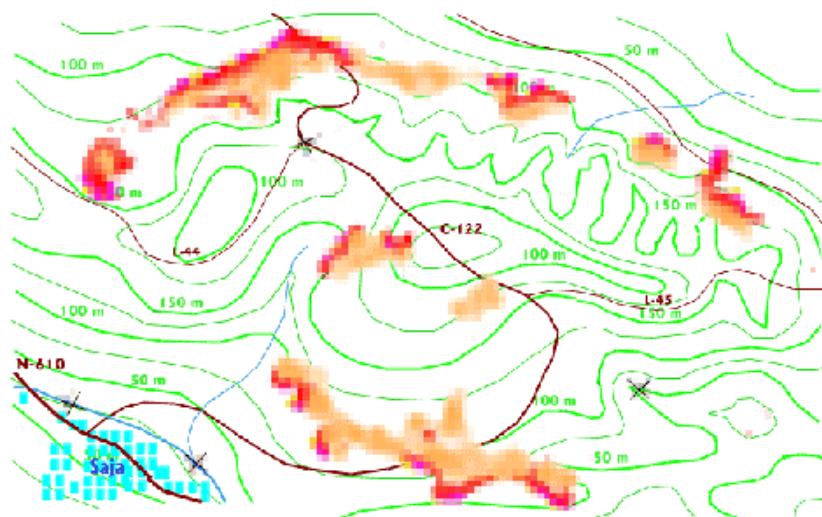


Fig. 89: Ejemplo de producto FUEGO de monitorización

El sistema FUEGO puede proporcionar gran cantidad de información sobre características físicas de los sucesos de incendios forestales, pero parece difícil su uso para la prevención, sobre todo desde el punto de vista del pronóstico de comportamientos sociales, una de las mayores causas del

inicio de incendios forestales. Suele resultar complicado establecer una comunicación fluida entre un sistema de gran alcance y alta tecnología con los usuarios “a pie de fuego”.

Parece adecuada una interpretación intermedia de los datos suministrados por el sistema Fuego, con el fin de acercar cognitivamente los resultados a los usuarios finales. Las características específicas de los medios de lucha contra el fuego en una zona determinada juegan un papel fundamental en la capacidad de lucha contra los incendios y deben ser tenidas en cuenta a la hora de suministrar consejos en las tareas de prevención y actuación. Por otra parte el considerar el viento como factor de riesgo principal para la definición de las áreas de cobertura, parece adecuado una vez se ha declarado el incendio, en cuanto a la simulación del avance del mismo, y en determinadas localizaciones geográficas, pero no cuando se debe establecer un índice de peligro, cuyo alto valor puede estar motivado, por ejemplo, por cierta agitación social en una determinada área.

3. LATUV.

El trabajo desarrollado por el Laboratorio de Teledetección de la Universidad de Valladolid se basa en la utilización de los satélites de la serie NOAA, con una resolución de 1km. por pixel en el nadir, apropiada para la detección y evaluación de áreas afectadas por los incendios forestales en España. Mediante la combinación adecuada de diversas bandas puede determinarse el estado de la vegetación y elaborar distintos modelos para la determinación del Riesgo de Peligrosidad, delimitando así zonas de Riesgo. Una representación muy útil consiste en dividir todo el territorio en celdas de 10x10 km. y asociar un índice de riesgo a la celda. Durante la temporada del verano, el Laboratorio de Teledetección de la Universidad de Valladolid, procesa diariamente y en tiempo real esta información, a la que se tiene acceso desde Internet, por encargo del Área de Defensa contra Incendios Forestales de la Dirección General de Conservación de la Naturaleza del M.A.P.A.

De acuerdo con lo expuesto, el riesgo de Incendios Forestales se calcula en celdas de 10x10 km., correspondientes a las áreas donde el ICONA sitúa las cuadrillas de vigilancia. Para el cálculo se tiene en cuenta la evolución del NDVI durante épocas anteriores comparado con la imagen diaria capturada del Satélite NOAA-14. Aquí se muestra un mapa de riesgo diario para todas las comunidades autónomas (figura 96), sólo hay que pinchar en el mapa sobre la zona deseada para ver el mapa de riesgo diario correspondiente. Los mapas de riesgo son válidos para celdas forestales, y los números que hay dentro de las celdillas se corresponden a los siguientes índices de riesgo: 0 Sin riesgo, 1 Riesgo bajo, 2 Riesgo moderado, 3 Riesgo alto, 4 Riesgo extremo, + Sin datos

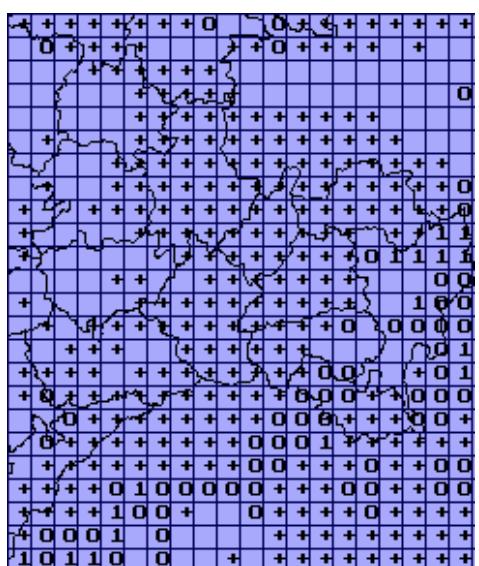


Figura 89

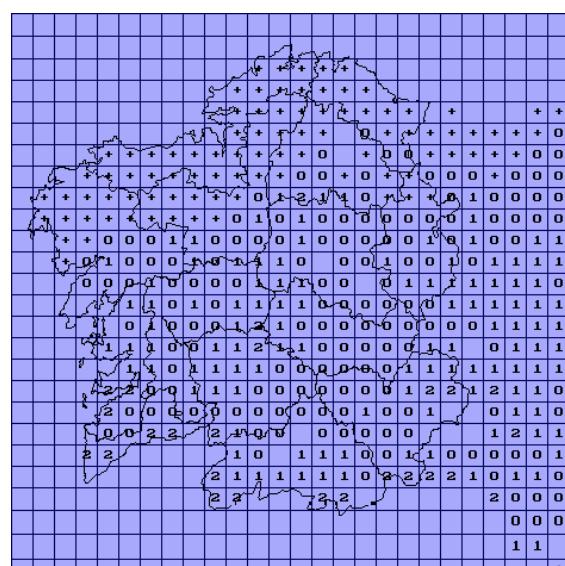


Figura 90: Ejemplo de Galicia. 26/3/98

Aunque muy útil, el índice de peligro suministrado por LATUV aparece como demasiado general, y se podría cuestionar la utilidad de esta información para los posibles usuarios finales. El tamaño de celda de 10x10 km. puede resultar útil, en regiones con mayores superficies de vegetación homogénea, como Andalucía o Castilla - La Mancha, pero parece adecuada una mayor granularidad en regiones como las de la cornisa cantábrica, en las que en una celda de este tamaño se puede encontrar una diversidad que haga necesario el tratamiento por separado.

Por último, reseñar los trabajos realizados por el grupo del Departamento de Econometría y Métodos Cuantitativos de la Facultad de Ciencias Económicas y Empresariales de la Universidad de Santiago de Compostela. Estos trabajos están enmarcados en la Teoría de Juegos de Estrategia. Los agentes considerados (Administración Central, Autonómica, Diputaciones, Ayuntamientos, particulares, etc.) están razonablemente obligados a cooperar, y se busca entonces un reparto equitativo de los frutos de la cooperación. En el modelo propuesto, se definen primeramente unas comarcas caracterizadas por un cierto grado de "homogeneidad" entre los puntos que las componen, para luego, utilizando técnicas semejantes a las del análisis imput-output evaluar comparativamente el nivel de "productividad" de los puntos estudiados. A continuación se introduce un "vector de precios" que permite comparar unas zonas con otras no sólo desde el punto de vista de riesgo de incendio sino también teniendo en cuenta su valor ecológico, social, etc. [Quiñoá 97].

En el modelo propuesto, se denomina "punto negro" a una masa forestal concentrada en un lugar geográfico y dotada de cierta homogeneidad, susceptible de ser afectada por un incendio. Dado un punto negro, se entiende por atributo cada una de las propiedades o cualidades del mismo, incluida la adecuación de cada uno de los posibles medios de prevención y extinción de incendios.

Considerando para simplificar el manejo que los atributos se refieren exclusivamente a medios de extinción de tres tipos (aéreos, mecánicos y humanos), un modelo matricial de distancias permite "evaluar" el grado de coincidencia de intereses de los distintos puntos respecto a los atributos señalados. El resultado alcanzado se resume en el siguiente teorema:

Sea $A = (\alpha_{ij})$ una matriz positiva tal que

$$\sum_{i=1}^n \alpha_{ij} < 1, \forall j, 1 \leq j \leq n$$

Entonces, dado un vector Z (vector beneficios) estrictamente positivo, y un vector β positivo, existen un vector $p(Z) = (p_1, p_2, \dots, p_n)$ de precios estrictamente positivo y un vector positivo $X(Z)$ tales que $C X(Z) = \beta$, donde $C = (c_{ij})$ y $c_{ij} = p_i b_{ij}$.

Según los autores, este resultado puede permitir al planificador tener una idea precisa del comportamiento de los distintos puntos dentro del sistema. Dado un vector β (bien social deseado) se puede encontrar un vector de precios que produciría un beneficio, por pequeño que sea, en cada punto, y dado éste, se puede hallar un "nivel de actividad" del mismo. La comparación de ambos es un índice a tener en cuenta a la hora de asignar los recursos.

Apéndice 3

Otras Aplicaciones del Método.

A continuación, se muestran con detalle dos aplicaciones del método descrito en esta memoria y se citan algunas otras donde ha sido aplicado parcialmente.

1. Reconocimiento de la aparición de incendios forestales usando sensores de estado sólido.

En la lucha diaria contra los incendios forestales, la detección temprana adquiere especial relevancia. En condiciones de poca visibilidad (niebla...) o zonas agrestes y montañosas, donde la accesibilidad es complicada, y el uso de sistemas de Infrarrojos inviable, es apropiado el uso de Sistemas Multisensores de Estado Sólido [Olivas, 97a], de reducido tamaño, y bajo consumo, debido a su bajo precio y la facilidad de instalación de una gran cantidad de sensores. Para tener un grado de fiabilidad alto se deben usar sistemas multisensores, los cuales se gestionan utilizando el método descrito en este trabajo.

En esta aplicación se trabaja con un Multisensor de Óxido de Estaño, el cual está compuesto de 32 sensores, con distintos dopados, de Platino y Paladio. La tecnología utilizada para hacer estos es mediante Película Delgada (*Sputtering*). Por tanto cada sensor tiene diferente respuesta frente a iguales condiciones, de manera que reaccionan de forma diferente. Esta característica es la que se va a utilizar, como factores que determinen con qué prototipo borroso existe mayor compatibilidad. Por tanto los factores van a ser los valores de las resistencias de los sensores que van a variar con la detección de uno u otro gas.

Los Prototipos que se van a utilizar, son las distintas posibilidades de detección según la vegetación que se esté quemando, esto hace que se usen 4 prototipos, para caducífolios, perennes, arbustos, maleza, dado que el gas desprendido por unos o por otros es distinto. Los Prototipos han sido representados como números borrosos, los cuales permiten obtener un grado de pertenencia a cada prototipo a partir del resultado sensorial, al cual se le denomina concepto (este resultado está normalizado), con los cuales se puede establecer el centro de cada prototipo, representado por un número borroso triangular.

Al tener una respuesta sensorial se obtiene como resultado del proceso los diferentes grados de compatibilidad con cada uno de los prototipos, estableciendo por tanto el tipo de vegetación del incendio, lo cual indica las características de la zona que está siendo quemada. Además según el nivel de ese grado de compatibilidad, se puede hacer un estudio de la distancia a la que se encuentra, teniendo en cuenta características como la dirección y velocidad del viento.

Los sensores de Óxido de Estaño (SnO_2), responden a la aparición de diferentes tipos de gases variando su resistencia. Tomando 32 de estos sensores con diferente grado de dopado (Platino), se obtienen resultados diferentes para cada uno de los sensores, con un rango de valores que va desde $\text{k}\Omega$ a $\text{M}\Omega$. Debido a esta diferencia, los valores se normalizan, por lo que ante la presencia de

determinada concentración de partículas sólidas en el aire y considerando que todos los sensores están a la misma temperatura, se obtiene un resultado para un multisensor, mostrado en la fig. 91:

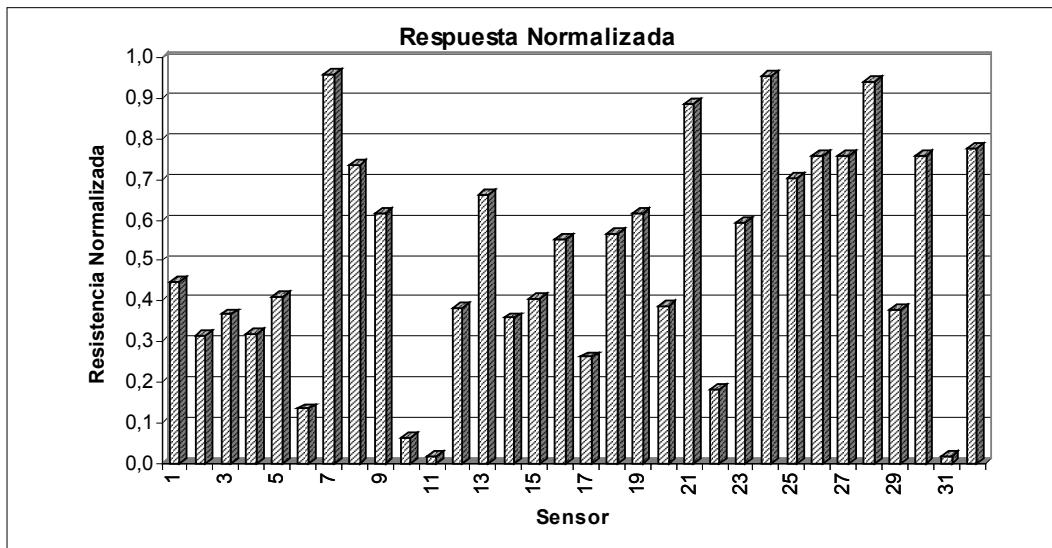


Figura 91: Resultado para un multisensor.

1.1. Determinación de los prototipos de las combustiones.

Los prototipos han sido representados como “números borrosos”, los cuales nos van a permitir obtener un grado de pertenencia a cada prototipo a partir del grado de pertenencia al concepto. En bien de la sencillez del modelo han sido representados como triángulos isósceles e iguales, aunque en determinadas condiciones se usan distribuciones trapezoidales, lo que no hace variar el método. Por tanto, para construir los prototipos (números borrosos triangulares) sólo necesitamos conocer su centro c_i (que puede ser un intervalo cuando son trapecios).

Estos Prototipos se obtienen mediante adquisición de conocimiento empírica, comprobando las respuestas del multisensor ante la combustión de diferentes materiales: maleza (P_1), arbustos (P_2), caducifolios (P_3) y perennes (P_4) y realizando un cálculo sobre los valores normalizados. Existe un umbral de disparo de alarma del multisensor, que viene dado por la mayor pertenencia a alguno de los prototipos y no al “estado normal”, (E.N.). Se reduce el error en el cálculo [Gardner, 92], utilizando como resistencia normalizada la siguiente expresión, Resistencia Fraccional:

$$R_F = (R_{\text{gas}} - R_{\text{aire}})/R_{\text{aire}}$$

Donde la resistencia del aire, se establece en el proceso de adquisición del conocimiento, y pertenece a la zona denominada E.N., este valor permanece constante para el procesado posterior.

Para calcular los centros de los prototipos c_i se utilizan pesos que multiplican a las Resistencias Fraccionales, de forma que se asegura una construcción de los prototipos como la representada en la fig. 99, sus centros serán:

$$c_i = \frac{\sum_{i=1}^{32} w_i R_{F_i}}{32}$$

y el cálculo de sus bases debe cumplir que $c_{i+1} - c_i < \text{Proyección ortogonal del lado del triángulo sobre el eje X}$, de esta forma se asegura el solapamiento, y se ratifica un correcto funcionamiento (fig. 92).

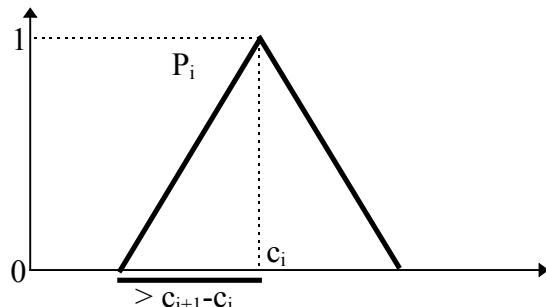


Figura 92.

Para encontrar el prototipo más afín a una respuesta real del multisensor se establece el siguiente proceso:

- #### I. Combinación de los valores generados por el multisensor en un caso real (r).

R_{Fi} : valor de cada sensor, normalizado. w_i iguales a los de los prototipos.

P_i : prototipo. $i=1$ para el prototipo “maleza”, etc.

$$r = \frac{\sum_{i=1}^{32} w_i R_{F_i}}{32}$$

- II. Calculo de los grados de compatibilidad con cada prototipo: $\mu P_1(r)$, $\mu P_2(r)$, $\mu P_3(r)$, $\mu P_4(r)$.

III. El prototipo más afín es el que mayor grado de compatibilidad tenga con r: un P_i tal que $\mu P_i(r) = \text{Max } \{\mu P_1(r), \mu P_2(r), \mu P_3(r), \mu P_4(r)\}$.

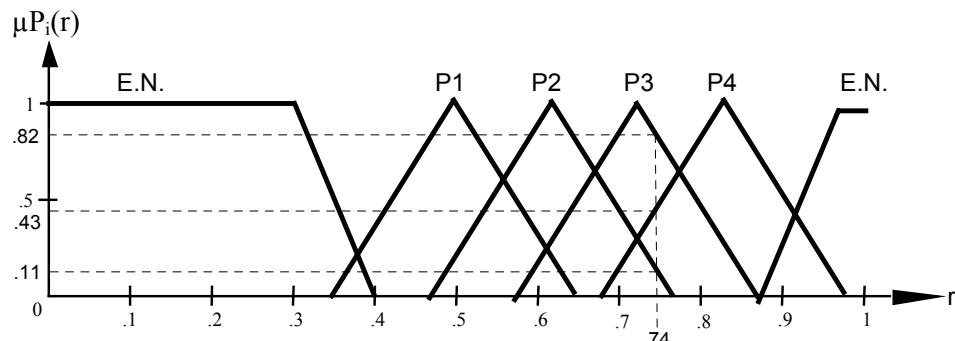


Figura 93: Ejemplo: Grados de Compatibilidad para $r = 0.74$

Como se puede observar en la fig. 100, los grados de compatibilidad son los siguientes:

$$\mu_{(E,N)} = 0, \mu P_1(r) = 0, \mu P_2(r) = 0.11, \mu P_3(r) = 0.82, \mu P_4(r) = 0.43$$

De lo que se deduce que el prototipo más afín es P₃ ("Caducifolios").

La obtención de un grado de afinidad con los prototipos preestablecidos proporciona información relevante para evaluar aspectos como dirección, intensidad y proximidad al incendio. En caso de existir una red de multisensores es viable evaluar con cierta precisión estas características, siempre que se disponga de datos adicionales como velocidad y dirección del viento, etc.

2. INFEDEC 2.1: Sistema de Ayuda en la Decisión de Diagnóstico y Tratamiento de enfermedades infecciosas.

Se presenta el sistema INFEDEC 2.1 [Lago, 96] como modelo de Sistema Basado en el Conocimiento aplicado al diagnóstico y tratamiento extrahospitalario en patología infecciosa de

vías respiratorias superiores y otorrinolaringología. INFEDEC no es un programa comercial y se encuentra a la libre disposición del personal sanitario.

El proyecto INFEDEC ha contado con la financiación del Fondo de Investigación Sanitaria (proyecto 94/1913), dentro del plan nacional de I+D, y se ha desarrollado con el apoyo de la Unidad de Investigación del Hospital Valle del Nalón (Asturias). Para comprobar el correcto funcionamiento del programa, se ha puesto a prueba en los siguientes centros sanitarios: Hospital "Valle del Nalón" (Asturias), Centro de Salud "La Calzada" (Asturias), Hospital de Caridad "Jove" (Asturias). El núcleo del proyecto INFEDEC es la realización de un sistema computacional que ayude a los médicos de atención primaria en estos dos aspectos:

- Gestión de la información (control de historiales clínicos, examen estadístico sobre epidemiología, etc.).
- Ayuda a la decisión médica, haciendo que la computadora simule el razonamiento y modo de solucionar los problemas de un profesional experimentado.

De esta forma el ordenador no es sólo un administrador de información, sino que además ayuda, mediante un conocimiento actualizado, en las decisiones clínicas. Otro aspecto importante es que de esta forma se consigue que los estudiantes en medicina realicen un aprendizaje interactivo, ya que pueden aprender mediante casos y obtener una respuesta razonada y justificada. Con INFEDEC se permite acceder a una enseñanza interactiva de estas enfermedades, ya que el ordenador permite contrastar decisiones diagnósticas y terapéuticas, además de asociar dichas decisiones con un texto descriptivo de la patología infecciosa. Las tareas fundamentales que realiza el programa aparecen en la fig. 94:

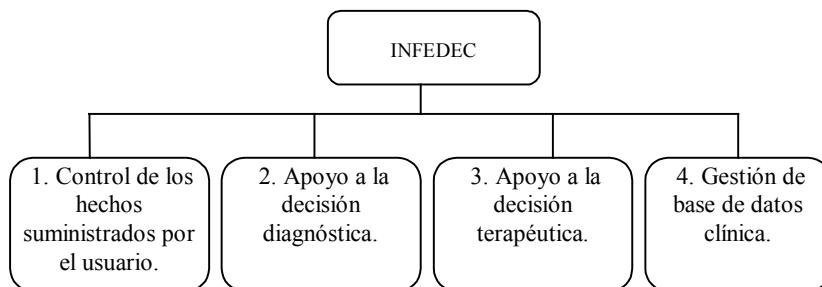


Figura 94: Tareas fundamentales del programa.

2.1. *Modelo lógico.*

Los pasos a seguir en el proceso de diagnóstico son básicamente los siguientes (fig. 95):

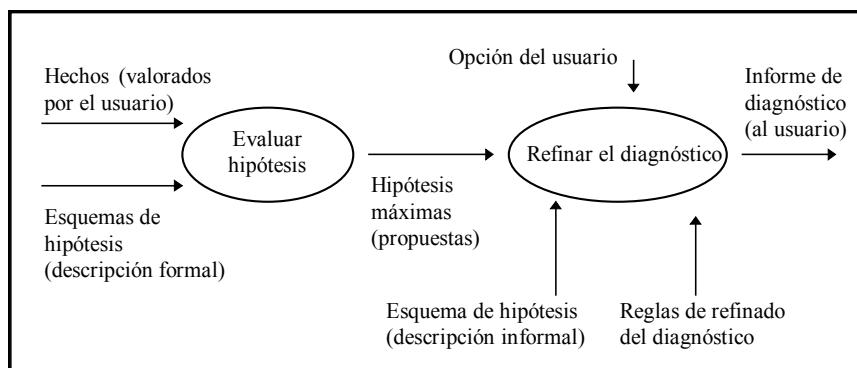


Figura 95: Proceso global.

- I. Evaluación de las hipótesis (prototipos de enfermedades): se obtiene una lista de hipótesis propuestas, ordenadas por su grado de credibilidad (grado de afinidad con los prototipos).

- II. Refinamiento del diagnóstico. Se seleccionan aquellas que han superado cierto umbral de aceptabilidad y se les añade la información textual (comentario, descripción, etc.). El usuario puede optar por una de las propuestas o por otra no presentada por el sistema.

El núcleo del proceso de diagnóstico se encuentra en el punto primero, por ello, nos centraremos en él. El razonamiento médico, al igual que el razonamiento científico y técnico sigue una lógica de conjeturas y refutaciones; es decir, el médico desde su encuentro con el paciente, comienza a emitir hipótesis diagnósticas que enseguida empiezan a ser contrastadas, estas hipótesis pueden ser verificadas o, por el contrario, ser refutadas. En el caso de ser refutadas, es decir, que no resistan la contrastación experimental; entonces son sustituidas por otras que se adapten mejor a los hechos. Este proceso no es esencialmente diferente del método de conjeturas y refutaciones expresado por Popper como un modelo de resolución de problemas. En este programa se representa este proceso asignando a cada hipótesis:

- Grado de confirmación, que nos indica la medida en que los hechos (o también otras hipótesis) apoyan la hipótesis que se está examinando.
- Grado de falsación, que representa la medida en que los hechos (o también otras hipótesis) refutan la hipótesis.

El resultado de combinar ambos factores es el grado de credibilidad de dicha hipótesis. Por lo tanto cada hipótesis representada por el programa incluye, entre otras cosas, un conjunto de hechos que la falsifican o confirman. El modelo lógico para representar los factores de confirmación, falsación y credibilidad ha sido el presentado en este trabajo. Se definen los prototipos de cada enfermedad y se comparan los síntomas con ellos. El grado de afinidad dará el grado de confirmación de cada una de las enfermedades.

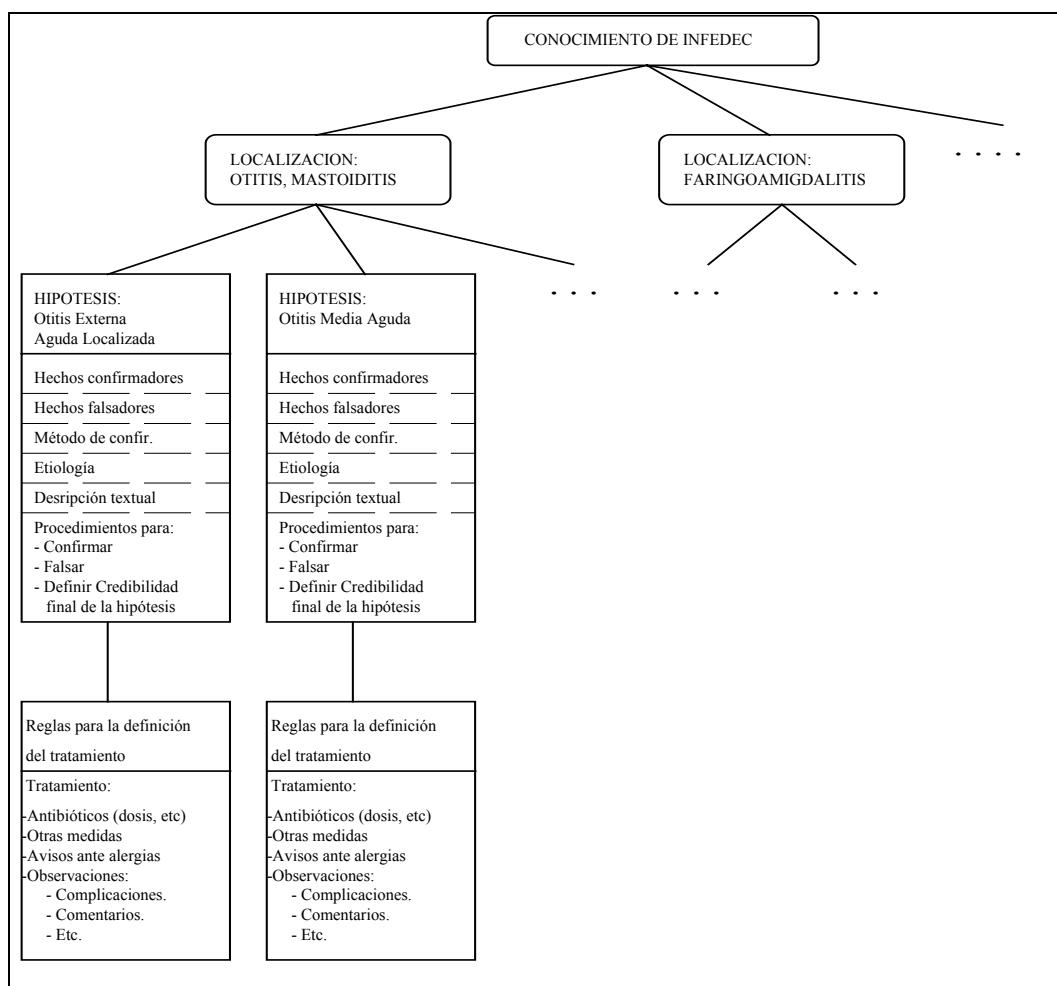


Figura 96: Conocimiento en INFEDEC.

2.2. Funcionalidades de INFEDEC.

FILIACIÓN DEL PACIENTE: INFEDEC permite introducir los datos de filiación del paciente que luego pueden ser almacenados en la base de datos.

DIAGNÓSTICO: A partir de los hechos introducidos por el usuario, INFEDEC presenta unas propuestas de diagnóstico de sospecha, ordenadas por su grado de credibilidad. Cada propuesta de INFEDEC está justificada, ya que se informa de las razones (hechos) que justifican dicha propuesta y se hace referencia a bibliografía. Además cada propuesta tiene una descripción textual que ayuda al usuario a tener una visión más completa tanto del diagnóstico como del tratamiento de dicha infección. Además informa de los microorganismos que pueden producir la infección, así como del método(s) de confirmación del diagnóstico. Se permite al usuario optar por una alternativa que no hubiera sido propuesta por el programa. Un ejemplo de propuestas de diagnóstico aparece en la fig. 97:

Propuestas de diagnóstico

PROPUESTAS DE DIAGNÓSTICO (Credibilidad)

Inf. del Tracto Urinario Inferior (Muy alta)
Inf. del Tracto Urinario Superior (Baja)

HECHOS ASOCIADOS CON EL PROCESO QUE ESTAN PRESENTES

Polaquíuria. Disuria. Dolor suprapúbico. Fiebre (37º-38º). Nitritos.

HECHOS ASOCIADOS CON EL PROCESO QUE NO ESTAN PRESENTES

Leucocituria.

AVISOS

DATOS ESPECIALMENTE RELEVANTES QUE SON DESCONOCIDOS:

Leucocituria.

COMENTARIO

En mujeres jóvenes no embarazadas y sin factores de riesgo podría realizarse tratamiento sin cultivo previo ni posterior
[STAMM, 1993].

Confirmación Microorganismos Descripción Alternativa Cerrar

Figura 97: Ejemplo de propuestas de diagnóstico.

TRATAMIENTO: A partir del diagnóstico INFEDEC presenta unas propuestas de tratamiento empírico, ordenadas por su grado de adecuación (preferencia). Para cada propuesta se especifica las características de administración, sin presentar marcas comerciales. Además se muestra otras posibles medidas, así como una serie de observaciones. El usuario puede anotar una alternativa que no haya sido propuesta por el programa. Un ejemplo se puede ver en la fig. 98:

Propuesta(s) de tratamiento			
JUSTIFICACION DE LA(S) PROPUESTA(S)			
Diagnóstico: Inf. del Tracto Urinario Inferior			
MEDIDAS FARMACOLÓGICAS			
Trimetoprín/Sulfametovazol Amoxicilina/Clavulánico	Orden de elección: 1	Vía: vo	Dosis: 500
Norfloxacina Doxiciclina Fosfomicina Trometamol	Duración: 3 días	Intervalo: 12	
OTRAS MEDIDAS			
En caso de reinfecciones frecuentes estaría indicado un tto profiláctico con el objetivo de mantener al paciente libre de infección urinaria al menos 6-12 meses consecutivos [REESE 86]; también micción postcoital en mujeres jóvenes [STAMM 1993].			
OBSERVACIONES			
El TMP/SMX es universalmente aceptado como antibiótico de elección en las infecciones urinarias por su actividad, modo de acción, comodidad de administración y precio. Si bien cualquier quinolona podría estar indicada en ITU, se reserva			
ALTERNATIVA: _____			
Cerrar			

Figura 98: Ejemplo de tratamiento.

VADEMÉCUM: Se presenta un vademécum adaptado al ámbito del programa (infecciones respiratorias y del tracto urinario).

BIBLIOGRAFÍA: Se ofrece también una lista de referencias bibliográficas utilizadas.

BASE DE DATOS: Una vez que se haya(n) realizado la(s) decisión(es), se pueden salvar en la base de datos de actos asistenciales, incluyendo datos de filiación del paciente. Sobre la base de datos se pueden realizar consultas. A partir de la base de datos el programa nos muestra la distribución estadística de los pacientes de acuerdo con dos factores: sexo y edad.

REQUERIMIENTOS HARDWARE: Los requisitos hardware son un ordenador 386 o superior con al menos 2 Mb. de RAM y 5 de Disco Duro. INFEDEC se ejecuta sobre sistema Windows 3.0 o superior.

3. Otras aplicaciones.

Como muestra de otras aplicaciones llevadas a cabo con parte de las ideas presentadas en este trabajo (metodología, adquisición de conocimiento, predicción, marcos ...), se citan algunos de los proyectos fin de carrera dirigidos por el autor, en los que estos mecanismos se han validado y exportado a otros problemas y dominios del mundo real, incluidos los sistemas distribuidos e Internet, objetivo central de muchos de los campos de investigación actuales:

1. "SECH: Sistema Experto de Confección de Horarios". Lucía López San Miguel. Universidad Pontificia de Comillas. Escuela Técnica Superior de Ingeniería (ICAI). Ingeniero en Informática. Premio Extraordinario. 1998.
2. "SS. EE. Para la evaluación de la aplicación de una metodología de desarrollo de software". Angela María Conde Fernández. Universidad Pontificia de Comillas. Escuela Técnica Superior de Ingeniería (ICAI). Ingeniero Técnico en Informática de Gestión. 1998.
3. "Sistemas Distribuidos: Aplicación a un Sistema Basado en el Conocimiento". Pilar Aránzazu Ocaña Díaz-Ufano. Universidad Pontificia de Comillas. Escuela Técnica Superior de Ingeniería (ICAI). Ingeniero en Informática. 1999.

4. “WINCREDIT: Sistema Inteligente de Ayuda a la Decisión en la Concesión de Créditos Bancarios”. Valentín Cabello Blanco. Universidad Pontificia de Comillas. Escuela Técnica Superior de Ingeniería (ICAI). Ingeniero en Informática. 1999.
5. “Sistema Experto de Ayuda a la Decisión para riego y abonado de una finca agrícola”. Pedro Miguel Martínez Trigueros y M^a Angeles Vázquez Fernández. Universidad Pontificia de Comillas. Escuela Técnica Superior de Ingeniería (ICAI). Ingeniero Técnico en Informática de Gestión. 1999.
6. “Sistema Experto de Ajedrez”. Jaime Valero de Bernabé y Juan Santiago Rubio. Universidad Pontificia de Comillas. Escuela Técnica Superior de Ingeniería (ICAI). Ingeniero Técnico en Informática de Gestión. 1999.
7. “Sistema Experto de Técnicas y Tácticas de Fútbol”. Julio Lerena Fernández. Universidad Pontificia de Comillas. Escuela Técnica Superior de Ingeniería (ICAI). Ingeniero Técnico en Informática de Gestión. 1999.
8. “Gestión de una Agencia de Viajes mediante un Sistema Experto”. Juan Antonio López Fernández-Cañadas. Universidad Pontificia de Comillas. Escuela Técnica Superior de Ingeniería (ICAI). Ingeniero Técnico en Informática de Gestión. 1999.
9. “Sistema Inteligente de Reconocimiento de Caracteres”. Elisa Martín. Universidad Pontificia de Comillas. Escuela Técnica Superior de Ingeniería (ICAI). Ingeniero Técnico en Informática de Gestión. 1999.
10. “Sistema Experto de Diagnóstico Odontológico (SEDO)”. Gonzalo Vázquez Pérez. Universidad Pontificia de Comillas. Escuela Técnica Superior de Ingeniería (ICAI). Ingeniero en Informática. 1999.
11. “Sistema Experto Monitor de Adiestramiento Deportivo”. Manuel López García. Universidad Pontificia de Comillas. Escuela Técnica Superior de Ingeniería (ICAI). Ingeniero Técnico en Informática de Gestión. 2000.
12. “Optimización Inteligente del Método *Stepping-Stone* de Distribución de Recursos”. Javier Barrachina García. Universidad Pontificia de Comillas. Escuela Técnica Superior de Ingeniería (ICAI). Ingeniero en Informática. 2000.