

3 Razonador Basado en Casos

El Razonamiento Basado en Casos (CBR) es una rama de la inteligencia artificial que se preocupa por el estudio de los mecanismos mentales necesarios para repetir lo que se ha hecho o vivido con anterioridad, ya sea por uno mismo, o ya sea por casos concretos recopilados en la bibliografía o en la sabiduría popular. Los diversos casos son del tipo "Si X, entonces Y" con algunas adaptaciones y críticas según las experiencias previas en el resultado de cada una de dichas reglas .

Casi no es posible imaginar un campo de mayor trascendencia para el humano que el de la toma de decisiones. Tenemos un problema cuando no sabemos cómo seguir. Una vez que tenemos un problema, hay que tomar una decisión (incluyendo la de no hacer nada). Elegimos una alternativa que nos parezca suficientemente racional que nos permita más o menos maximizar el valor esperado luego de resuelta nuestra acción. Emitimos en silencio un plan de control, que nos guía en la toma de decisiones, incluyendo decisiones relacionadas con modificar ese plan de control. Demostramos nuestra inteligencia en este proceso en funcionamiento.

Un tópico bastante poco conocido es el papel de la inteligencia en la resolución de problemas, de la toma de decisiones, del planeamiento de un control. Para la realización de un CBR debe de cumplirse ciertos lineamientos y comprender su estructura.

3.1 Estructura

La estructura de este tipo de sistemas se divide en cuatro pasos (Figura 3.1) :

1. Recuperar (retrieve) los casos más parecidos (un nuevo problema se aparea con casos similares guardados en la base de casos)
2. Reutiliza (reuse) la solución propuesta en los casos para tratar de resolver el problema
3. Revisar (revise) la solución propuesta (en caso de ser necesario)
4. Almacenar (retain) la nueva solución como parte de un nuevo caso

Un caso es un "pedazo" de conocimiento contextualizado representando una experiencia. Los casos se encuentran formados por:

1. el problema que describe el estado del mundo cuando ocurrió el caso
2. una descripción de la solución encontrada y/o
3. un resultado describiendo el estado del mundo después de que ocurrió el caso

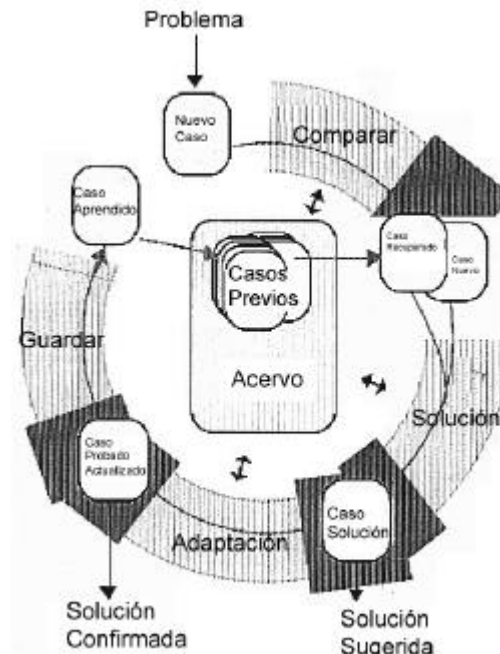


Figura 3.1 Estructura de un CBR [AAMODT 94]

Los casos pueden representarse de diversas formas, pero la representación frames/objetos es la más usada.

Para facilitar la recuperación asignar índices. Los índices deben :

- ?? ser predictivos
- ?? contemplar el propósito para el cual el caso se va a utilizar
- ?? ser suficientemente abstractos para ensanchar el uso de la base de casos
- ?? ser suficientemente concretos para ser reconocidos en el futuro

Dentro de los métodos de indexación automáticos están [Morales 99]:

- ?? Indexación por atributos o dimensiones que son responsables de la solución o que influyen en el resultado (checklist)
- ?? Indexación basada en diferencias: selecciona por índices que diferencian un caso de otro
- ?? Métodos basados en similitudes o en explicaciones: produce un conjunto de índices para casos abstractos creados a partir de casos que comparten atributos comunes. Los atributos que no se clasifiquen en su similitud y que no generen índices, se utilizan como índices de los casos generales.
- ?? Métodos inductivos: identifica atributos predictivos y los utiliza como índices (variantes de ID3)

?? Técnicas basadas en explicaciones: determina los atributos relevantes a cada caso e indexarlos por medio de ellos

El problema de encontrar el “mejor” caso se ha tratado en analogía. Esto involucra heurísticas para restringir/dirigir la búsqueda. El razonamiento basado en casos servirá para problemas de gran escala cuando se resuelvan los problemas de recuperación eficiente con miles de casos

Las heurísticas deben de permitir hacer apareos parciales

Métodos:

?? Vecinos más cercanos: suma pesada de atributos que aparecen. Problema: como determinar los pesos. Algoritmo típico [Morales 99]:

$$\frac{\sum_{i=1}^n w_i \times sim(f_i^I, f_i^R)}{\sum_{i=1}^n w_i}$$

donde:

w_i es el peso de importancia de un atributo

sim es la función de similitud

f_i^I y f_i^R son los valores del atributo i en el caso de entrada (I) y el caso recuperado (R)

- ?? Inducción determina que atributos mejor discriminan casos y genera un árbol de decisión para organizar los casos en memoria
- ?? Basado en conocimiento: aplica conocimiento del dominio (no necesariamente completo) para identificar los atributos
- ?? Por medio de templates: parecido a queries tipo SQL, recupera todos los casos que cumplen con ciertos parámetros (muchas veces se usa antes de las otras técnicas)

Una vez que se recupera un caso, la solución debe de adaptarse. La adaptación se fija en las diferencias entre los casos y aplica reglas de adaptación. La adaptación puede ser:

- ?? Estructural: la adaptación con reglas es sobre la solución
- ?? Derivacional: reutiliza los algoritmos, métodos o reglas usados para generar la solución (la solución debe de guardarse con la secuencia o plan utilizado). También se conoce como reinstanciación

Tipos de adaptación:

- ?? Nula (útil con problemas complejos con soluciones simples)
- ?? Ajuste de parámetros (compara parámetros específicos)
- ?? Abstracción y re-especialización

- ?? Adaptación basada en críticas (combinaciones de atributos)
- ?? Reinstanciación (instanciación de atributos usados)
- ?? Volver a derivar la solución
- ?? Reparación guiada por modelos
- ?? Substitución basada en casos (usar CBR para sugerir adaptaciones)

3.2 Trabajos Relacionados

Dentro de este tema se encuentran trabajos que pueden ayudarnos para tomar mejores ideas para este proyecto que se esta elaborando entre ellos se encuentra:

- ?? A Case-Based Urban Planning Support System Using An Integrated Combination of Geographical Information Systems and Remote Sensing, desarrollado por Ko-Wan Tsou, Yao-Lin Chang, Yu -Ting Hung[TSOU01]
- ?? Active DSS for industrial Emergency Management Based on GIS, desarrollado por Giovanni Di Constanzo y Adam M. Gadomski[CONSTANZO98]
- ?? Case-Based Reasoning and Learning Agent, desarrollado en la Universidad Politécnica de Cataluña [Sánchez-Marrè 97]

A continuación se dará una breve descripción de lo que se desarrolló en estos trabajos.

3.2.1 Active DSS for industrial Emergency Management Based on GIS

Este trabajo es desarrollado por Protección Civil de Italia y el Sistema de Soporte de Decisión [Figura 3.2]. El proyecto esta enfocado para prevenir perdidas en intervenciones en tiempo-real y reducir posibles pérdidas en población y el medio ambiente. Este sistema esta compuesto de tres módulos: Diagnóstico, Predicción y Toma de decisión. Actualmente el trabajo esta concentrado en identificar procesos de sustancias tóxicas tanto sus posibles causas como reporte de síntomas encontrados.

Diagnóstico

Se enfoca a reconocer el accidente que ocurre, usando los síntomas y usando el método del CBR. Como primera respuesta el sistema presenta una posible causa, las industrias cercanas y sustancias. El operador del sistema introduce la descripción del caso

en la computadora y el sistema a través de un razonador predefine casos en base a la experiencia que tenga el sistema en el, y sugiere soluciones al problema.

Predicción

Este módulo realiza preguntas para prever posibles evoluciones de escenarios y asume que el caso viene del módulo de diagnóstico, lo que ayuda a prevenir ciertas características que no son tomadas en ese momento de emergencia. Su meta es indicar la máxima consecuencia negativa sobre el evento, así como también genera posibles escenarios de emergencia que pueden ser consecuencia del caso analizado.

Toma de decisión

toma como entrada los datos que son consecuencia del accidente, identifica los pasos previos y aplica los procedimientos de intervención en las situaciones de emergencia, en algunos casos preventivos. La salida del sistema consiste en acciones o grupos de acciones que el sistema sugiere como respuesta a los eventos del accidente presentado.

Un caso se encuentra representado por:

- ?? Título del problema que el caso representa
- ?? Descripción del problema
- ?? Un conjunto de atributos representado por parámetros o variables del propio caso
- ?? Un conjunto de sugerencia o acciones a ser ejecutadas para solucionar el caso

Los métodos para encontrar casos similares al caso actual son:

- ?? comparación textual entre los casos descriptivos, nivel de palabras, conjunto de caracteres o frases, con ayuda de un vocabulario de sinónimos asociado con el Caso Base
- ?? Análisis de valores ejecutados en un atributo

El sistema cuenta los siguientes requerimientos básicos:

- ?? Activación manual en base a el registro de síntomas
- ?? Soporte en el reconocimiento de sustancias tóxicas
- ?? Soporte en la identificación de la industria en la cual surge la emergencia

- ?? Soporte en el manejo de intervenciones adecuadas que son reconocidas por la situación que fue analizada

El sistema usa dos bases de datos diferentes:

- ?? Una base de datos de riesgos industriales
- ?? Una base de datos de sustancias tóxicas, con asociación a síntomas

3.2.2 A Case-Based Urban Planning Support System Using An Integrated Combination of Geographical Information Systems and Remote Sensing

Es un sistema de planeación urbana trabaja con CBSS [Figura 3.2] (Case-based support system) que es una innovación a los KBS (knowledge-based system). Evalúa casos históricos para comparar con planes espaciales y el uso reciente de la tierra.

Cuenta con una librería de casos, la característica de los casos de planeación son categorizadas de la siguiente forma:

- ?? Describe de los atributos de información de características espaciales
- ?? Describe información espacial y el grado de desarrollo urbano obtenido de la relatividad de los objetos
- ?? Describe atributos de información no espacial de características que expresan el significado de la planeación
- ?? Describe la relación de características no espaciales para elaborar relaciones no espaciales con otras características

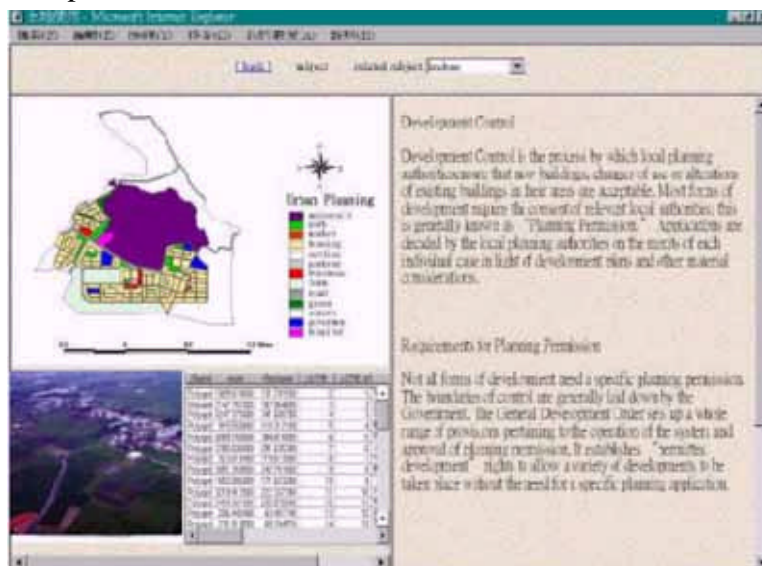


Figura 3.2 Prototipo CBUPSS

3.2.3 Case-Based Reasoning and Learning Agent

El sistema cuenta con cuatro tipo de archivos de datos :

- A) El archivo de definición de variables (que describe cada variable de que tipo es, que intervalos se definen)
- B) El archivo de prioridades entre las variables (da más peso a unas variables respecto de otros)
- C) El archivo de datos que se pueden dar como entrada al CBR
- D) El archivo que guarda toda la estructura de la librería de casos en disco.

Para que el sistema funcione deben de realizarse los siguiente pasos:

- crear una librería de casos
- añadir casos a la librería se puede hacer a partir de pantalla en la cual va preguntando los valores para llenar la ficha del caso o bien desde un archivo con la información y estructura específica
- una vez acumulada la experiencia su puede buscar el caso similar mediante un proceso de manejo de distancias en la cual también se cuenta con un menú de opciones para escoger el tipo de distancia que se ocupara en el resto del sistema.

Este sistema se encuentra desarrollado en COMMON LISP ALLEGRO:

- ?? cuenta con una librería de casos que lleva una estructura específica,
- ?? cuenta con un menú de opciones donde el usuario va desarrollando sus casos, es decir personaliza el tomador de decisiones,
- ?? cuando comienza el sistema a trabajar, el usuario define las variables que entraran a ser evaluadas, les da un peso y un grado de valor que pueden obtener es decir el atributo puede ser medido en base a una escala numérica por decir de 1...10 o puede ser evaluada como etiquetado bajo, mediano, alto.

El peso que se le da es para que al momento de buscar el caso más cercano se toman en prioridad atributos de semejanza. Para la búsqueda de los casos cercanos a él, el usuario puede también determinar el tipo de evaluación de dirección que quiere que el sistema tome en cuenta en el momento de realizar el calculo de distancias para obtener el caso más cercano. También se tiene la opción de realizar un desplegado de la librería de casos y sus atributos que se encuentran interactuando. Este sistema es un evaluador, mas no un adaptador de casos, es decir, que si se quiere hacer una mejora al caso esto no es posible de realizar, este sistema lo único que realiza es la evaluación del caso y soluciones que pueden ser tomadas para resolverlo, en base a casos previos.

Trabaja con una estructura jerárquica de la librería de casos, y ello lo hace en base a los atributos significativos del caso, el cálculo de los casos cercanos lo hace en base a la distancia. El usuario tiene una gran variedad de métodos para realizarlo y debe ser definido al iniciar el sistema. Trabaja directamente con archivos planos y no tiene acceso a base de datos. Los atributos de la ficha del caso son del tipo par Atributo-Valor. Estos tipos de datos son del tipo:

?? Categórico : Ejem. Tamaño, peso, distancia, metros, etc.

?? Lineal :

- o Ordenado: Ejem. Tamaño {grande, mediano, pequeño}
- o No ordenado : Ejem. Color {Rojo, Verde, Azul}

Veamos un ejemplos para dar una idea mas clara de lo que se ha dicho, supongamos que queremos hacer una librería de casos para el volcán donde los atributos que se definan son del siguiente tipo:

Nombre del atributo: ZONA
Tipo: CATEGORIC
Numero de modalidades: 3
Modalidad numérica: 1
Modalidad cualitativa: P_MAYOR
Modalidad numérica: 2
Modalidad cualitativa: P_MODERADO
Modalidad numérica: 3
Modalidad cualitativa: P_MENOR
Peso global del atributo (0..10): 10.0

Nombre del atributo : MATERIALES
Tipo : LINEAL
Numero de modalidades: 3
Modalidad numérica: 1
Valor inferior: 10.0
Valor superior: 30.0
Modalidad cualitativa: MAYOR
Modalidad numérica: 2
Valor inferior: 3.0
Valor superior: 9.0
Modalidad cualitativa: MENOR
Modalidad numérica: 3
Valor inferior: 1.0
Valor superior: 2.0
Modalidad cualitativa: MODERADA
Peso del atributo (0..10): 7.0

Nombre del atributo: FLUJO
Tipo: LINEAL
Numero de modalidades: 3
Modalidad numérica: 1
Valor inferior: 54.0
Valor superior: 80.0
Modalidad cualitativa: ALTA
Modalidad numérica: 2
Valor inferior: 30.0
Valor superior: 53.0
Modalidad cualitativa: MEDIA
Modalidad numérica: 3
Valor inferior: 1.0
Valor superior: 29.0
Modalidad cualitativa: BAJA
Peso global del atributo (0..10): 6.0

Se alimenta el sistema con casos y el árbol

La estructura del árbol que genera es la siguiente:

1-ZONA

1-P_MAYOR fr: 2

(ZONA: P_MAYOR ↯ Zona en la que se ubica el caso
MATERIALES: 6.0 ↯ Cantidad de material registrado
FLUJO: 8.0 ↯ Cantidad de flujo registrado
dist: 0.038793027
id: C3)

1-P_MENOR fr: 1

2-FLUJO

2-MEDIA fr: 1

(ZONA: P_MENOR
MATERIALES: 4.0
FLUJO: 31.0
dist: 0.9832396
id: C1)

2-ALTA fr: 0

(ZONA: P_MENOR
MATERIALES: 15.0
FLUJO: 56.0
dist: -1.0
id: C2)

que mirar si lo que estamos buscando tiene $ZONA = P_MENOR$ o $ZONA = P_MAYOR$. En este caso es P_MENOR , así que nos vamos por la rama de abajo. El siguiente atributo que tenemos que mirar es flujo. Nos pregunta si flujo es $MEDIA$ o $ALTA$. Nosotros tenemos un número (55.0), ya que al definir las variables lineales se tiene que establecer unos rangos de valores, ahora vamos a recuperar la definición de la variable $FLUJO$ del que queda almacenada en el archivo `volcan.dat`. (el nombre `volcan.dat` es con referencia al ejemplo puede dársele otro nombre según el CBR que se este generando).

```
("FLUJO"  
 .#S(ATRIBUT-GLOBAL  
 :TIPUS "LINEAL"  
 :NMOD 3  
 :LMOD ((1 54 80 "ALTA") (2 30 53 "MEDIA") (3 1 29 "BAJA"))  
 :PES 6))
```

Como podemos ver en la línea que empieza por `:LMOD` están definidas las modalidades (es decir, los intervalos). Si miramos a que intervalo corresponde el valor 55.0 podemos ver que es "ALTA". Con esto elegimos de nuevo la rama inferior y llegamos a un nodo en el que hay un solo caso (C2), y por lo tanto no hay que preguntar más: C2 es el caso más similar al caso nuevo C4. Por lo tanto este caso se devolverá, calculando la distancia el caso actual (en este caso la distancia será muy pequeña, ya que varía solo un punto el valor del flujo).

Por último el campo *fr* mide la frecuencia de paso por el nodo de la estructura. Que sirve para poder decidir que hacer en caso de que el caso que entra tenga un valor nulo en uno de los atributos que le preguntan. Como no hay valor, lo que se hace es irse por la rama del árbol Discriminante que ha sido más frecuentemente visitada.

El campo *dist* que tienen todos los casos es la distancia calculada entre ese caso y el último caso con el que se ha comparado. Una distancia $=-1.0$ significa que no ha sido comparado aún con ningún otro caso (es el valor inicial). Al igual que en el caso de *fr*, no es relevante ver que valor tiene en un momento dado cuando analiza la estructura de casos. Solo es relevante cuando el sistema devuelve el caso como respuesta a una petición, dando la distancia entre el caso actual y el caso más similar que ha encontrado (como al final de ejemplo de C2 y C4).

3.3 GIS, CBR y Volcán

Como ya sabemos los riesgos naturales son difíciles de combatir, pero gracias a la tecnología que esta en constante evolución y el surgimiento de los GIS, nos permiten ayudar a tomar medidas de precaución para así mitigar desastres devastadores que pueden dejar a su paso una amenaza natural.

Todo éste proceso involucra el tener mucha experiencia y conocimiento, para así poder tomar una decisión precisa. Pero para que se llegue a tener esta experiencia para mucho tiempo, implica aprender de errores del pasado para así no cometerlos en el futuro y porque no mejorarla. Es aquí donde entran en gran apoyo el CBR. A pesar de que los GIS son una herramienta que ayuda a determinar posibles efectos en el terreno, no se cuenta dentro de ella un método que permita evaluar eventos pasados para dar así un pronóstico de solución.

En el presente trabajo se trata de introducir el concepto y metodología de el CBR a los GIS haciendo algunas adaptaciones originales debido a que el CBR trabaja mucho mediante reglas y los casos que se modelarán en éste proyecto son con base a información cartográfica y sobre eventos que pueden ocurrir en nuestro contexto principal – el Volcán Popocatepetl - y el cual no nos permite poder determinar reglas de comportamiento. Esto nos lleva a crear un sistema de Razonamiento Basado en Casos con la modalidad de que la información de casos se obtiene directamente de la cartografía.

Al concluir el trabajo se demostrará que la combinación de un CBR y un GIS ayudará de manera eficaz en el proceso de toma de decisión en el momento en que ocurra una amenaza dando como resultado opciones de acciones que pueden ser tomadas como medidas preventivas en el momento de los hechos.

3.4 Resumen

Con base a las características descritas para un CBR se tomarán las pautas para el diseño del proyecto que se propone para un Soporte en la toma de decisiones. Además que de los sistemas descritos se tomarán en cuenta el diseño del CBR, de manera que el

producto que se entregue sea de resultados satisfactorios ya apegados los lineamientos de un CBR. Con la diferencia de que los atributos que describan un caso serán obtenidos de la cartografía, los atributos que mas interactúan son las poblaciones y las rutas de evacuación por lo cual la propuesta de atributos descriptivos que se puede obtener de la cartografía en base a la zona de amenaza registrada es:

- ?? *Fecha* de ocurrencia del evento
- ?? *Núm. De Poblaciones* afectadas por la zona de amenaza
- ?? *Metros de la ruta1 ... ruta10* que se encuentran amenazadas
- ?? *Total de metros afectados* de las 10 rutas que forman la evacuación
- ?? *Factor de estimación* que dará el experto al evento que ocurre.