

Xavier Alamán Roldán

Instituto de Ingeniería del Conocimiento, UAM
Cantoblanco

Email: alaman at IIC.UAM.ES

Redes bayesianas como nivel superior de representación del conocimiento en sistemas inteligentes de control supervisor

Resumen: en este artículo se propone el formalismo de las redes bayesianas como mecanismo de representación del conocimiento en la capa superior de una arquitectura genérica para sistemas inteligentes de control supervisor. Dicha arquitectura ha sido descrita en anteriores artículos (ver [Alamán et al. 91, 92]), y ha sido probada en una aplicación industrial: el sistema MIP (Monitorización Inteligente de Procesos). MIP es un sistema inteligente en tiempo real desarrollado en el Instituto de Ingeniería del Conocimiento (IIC).

Abstract: in this paper the bayesian networks formalism is proposed as the knowledge representation for the higher-level layer of a general architecture for intelligent systems for real-time supervisory process control. This architecture has been described elsewhere [Alaman et al 92; 92b], and has been tested in a real industrial case: the MIP system. MIP is a real-time expert system developed at the IIC.

1. Introducción

En este artículo se propone el formalismo de las redes bayesianas como mecanismo de representación del conocimiento en la capa superior de una arquitectura genérica para sistemas inteligentes de control supervisor. Dicha arquitectura ha sido descrita en anteriores artículos (ver [Alamán et al. 91, 92]), y ha sido probada en una aplicación industrial: el sistema MIP (Monitorización Inteligente de Procesos). MIP es un sistema inteligente en tiempo real desarrollado en el Instituto de Ingeniería del Conocimiento (IIC). Dicho sistema monitoriza, diagnostica y emite recomendaciones en tiempo real acerca de tareas de optimización y mantenimiento de la estabilidad relativos a una planta petroquímica. MIP está en producción desde Marzo de 1991 en una planta de síntesis de acrilonitrilo de REPSOL Química en Tarragona, con gran efectividad desde el punto de vista técnico y de resultados económicos.

El trabajo presentado en este artículo se concentra en los aspectos técnicos que han sido propuestos dentro de la capa superior de dicha arquitectura genérica. El segundo apartado describe el problema a resolver. El tercer apartado presenta la arquitectura genérica propuesta para sistemas inteligentes de ayuda al control supervisor. El cuarto apartado presenta los motivos por los que se propone el formalismo de redes bayesianas para el nivel superior de dicha arquitectura. Finalmente el último apartado presenta los resultados obtenidos.

2. Identificación del problema

El objetivo principal de la arquitectura MIP es aportar una estructura genérica que permita el desarrollo de sistemas de Inteligencia Artificial para el control supervisor. Un sistema de control supervisor tiene como cometido la generación de avisos y recomendaciones que ayuden al diagnóstico de problemas y a la toma de decisiones realizadas por los operadores responsables del control de un proceso industrial. En la arquitectura MIP dichas recomendaciones atañen fundamentalmente a la optimización del proceso, mediante la

sintonización fina de los parámetros de la planta para optimizar su productividad, y a la mejora de la estabilidad de la planta, mediante el diagnóstico precoz de problemas que pudieran degenerar en problemas de estabilidad.

Los sistemas de control supervisor contribuyen al control de la planta esencialmente a través de la mejora en la eficiencia del trabajo realizado por los ingenieros de procesos y operadores de control. Aspectos que pueden ser mejorados gracias a la introducción de sistemas de control supervisor incluyen la complementación de los conocimientos de los operadores al darles acceso a parte del conocimiento que poseen los ingenieros de procesos, la unificación de las tácticas de control entre distintos operadores, el evitamiento de sobrecarga cognitiva de los operadores en caso de emergencia y durante la operación diaria, y la presentación a los operadores de información de alto nivel de abstracción.

El objetivo de la arquitectura MIP es la definición de un marco general que permita la integración de diversas técnicas de Inteligencia Artificial para la implementación de tales sistemas inteligentes de supervisión.

3. La arquitectura MIP

La arquitectura MIP se basa en el concepto de pizarra ([Eрман et al. 80],[Engelmore y Morgan 88]). La pizarra es una estructura de datos compleja que contiene en sí misma los mecanismos de control, consistencia y refresco que aseguran en todo momento la calidad y coherencia de la información en ella contenida. La característica esencial de la pizarra es que todos los módulos del sistema usan esta única estructura común de datos para el intercambio de información. Todo módulo que necesite datos elaborados por otro módulo distinto, los toma de la pizarra, volcando a su vez en ésta los resultados finalmente obtenidos. La pizarra es la única representación del conocimiento global común a todos los módulos del sistema. La pizarra es también el mecanismo fundamental empleado para la resolución colaborativa de problemas dentro de la arquitectura.

Por ejemplo (**figura 1**), el módulo de razonamiento local puede comenzar la resolución de un problema a partir del estado de la planta representado en la pizarra. Como resultado incorpora a la pizarra los diagnósticos locales del sistema que sean pertinentes. El módulo de razonamiento global se basará en este modelo para emitir un diagnóstico global de la planta, completando así la resolución del problema comenzada por el primer módulo. Otras fuentes de conocimiento podrían ser añadidas para complementar el diagnóstico de la planta. La estructura interna de la pizarra es jerárquica, representando los distintos niveles de abstracción existentes en el problema.

La base de la arquitectura está soportada por las mediciones realizadas sobre la planta. Dichas mediciones incluyen una serie de datos adicionales aparte del mero valor numérico de la variable del proceso, como pueden ser el periodo de validez

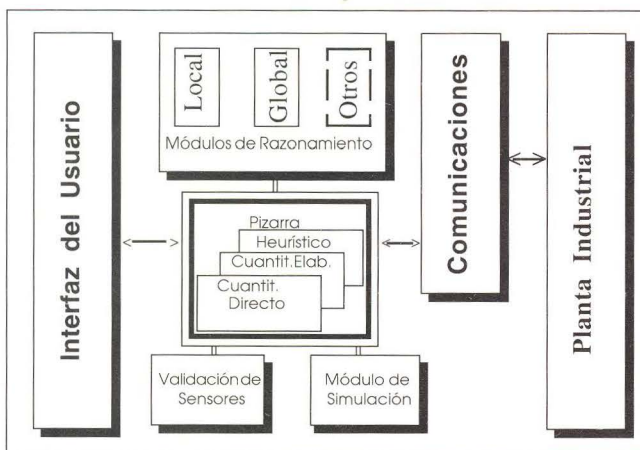


Figura 1: Arquitectura software del sistema MIP

del dato, las variables relacionadas, las fuentes alternativas del mismo, etc. Este nivel constituye el Modelo Cuantitativo Directo, que se concreta en forma de una serie de marcos y relaciones que modelan la estructura lógica y física de la planta.

Sobre el Modelo Cuantitativo Directo se establece un segundo nivel de conocimiento numérico y algorítmico (el Modelo Cuantitativo Elaborado). Gran parte del conocimiento que posee un experto en procesos industriales es expresable como un conjunto de modelos matemáticos. La implementación más natural de éstos es en forma de un simulador [Chang y Birdwell 90]. El Modelo Cuantitativo Elaborado puede ser complementado con otros aspectos de los datos. Uno de los más importantes entre dichos aspectos es la fiabilidad del dato obtenido. Dicha fiabilidad consiste en una representación de la incertidumbre (o credibilidad) de cada objeto de los niveles de conocimiento cuantitativos, que podrá ser proyectada posteriormente hacia niveles superiores.

El tercer nivel de conocimiento se formaliza como un sistema basado en reglas: el Modelo Heurístico Local. El conocimiento contenido en este nivel atañe a los diagnósticos locales del proceso. Por local entendemos el diagnóstico de posibles problemas en un subsistema hecho en base a variables cercanas a dicho subsistema desde un punto de vista lógico. La fuente de conocimiento de este nivel es el conjunto de reglas lógicas y heurísticas que el experto ha desarrollado a lo largo de su carrera profesional y que expresa aquellos aspectos locales del proceso que no pueden ser fácilmente modelados analíticamente, pero para los cuales el experto conoce soluciones y métodos heurísticos de diagnóstico [Efstathiou 85].

El cuarto nivel de conocimiento es un sistema de razonamiento bajo incertidumbre. Es el Modelo Heurístico Global, y contiene aquellos esquemas de razonamiento que permiten deducir problemas globales del sistema a partir de los diagnósticos locales del tercer nivel. Modeliza el esquema de razonamiento del experto, que primero obtiene conclusiones locales a partir del estado de la planta y luego combina e interrelaciona dichas conclusiones para obtener el diagnóstico global del estado del proceso. El diagnóstico global se realiza en base a las dependencias causales existentes entre los posibles problemas globales de la planta y sus manifestaciones en forma de disfunciones locales. Dichas dependencias se codifican en forma de una red bayesiana, como se describe en el siguiente apartado.

Una característica a destacar de esta arquitectura jerárquica es que cada nivel de conocimiento sólo emplea como primitivas

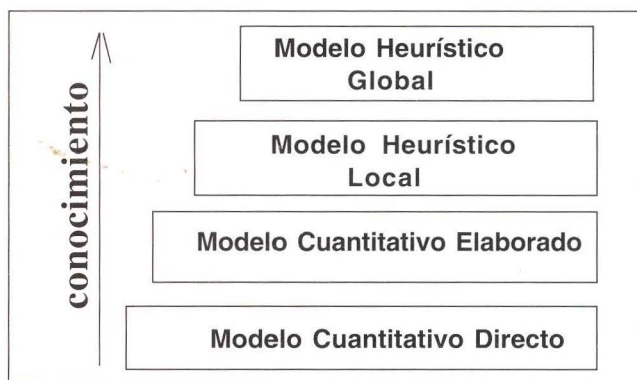


Figura 1: Arquitectura del conocimiento

aquellas conclusiones (o datos) que niveles inferiores han elaborado previamente. Este aspecto es esencial para una modularización adecuada de la arquitectura, además de ajustarse fielmente al esquema cognitivo del experto.

La arquitectura propuesta permite extender con facilidad la representación del conocimiento añadiendo nuevos niveles a los cuatro anteriores. Alguno de estos niveles podría ser implementado mediante formas alternativas de representación del conocimiento, como puede ser una red neuronal, una red semántica, etc., sin afectar sustancialmente al resto del sistema. De hecho, los distintos niveles de conocimiento pueden ser implementados como módulos separados dentro de una arquitectura de pizarra, lo cual permite su implementación e instalación completamente independiente.

El flujo de control entre las distintas capas de la pizarra suele ser de la siguiente manera. Hay objetos en la pizarra que representan variables de la planta (sensores), que tienen asociado un tiempo de refresco. Cuando el tiempo de refresco transcurre para una o varias variables, la pizarra manda una señal al módulo de comunicaciones, que se encarga de obtener nuevos valores para dichas variables. Cuando estos llegan, el módulo de simulación numérica y el módulo de validación de sensores reciben una señal, aportando entonces sus cálculos y razonamientos, y modificando donde fuera necesario el conocimiento presente en el nivel del Modelo Cuantitativo Elaborado. En el siguiente paso la mayor parte de las veces interviene el módulo de razonamiento local, que revisará y generará en caso necesario los diagnósticos adecuados acerca del estado de la planta. Finalmente el módulo de razonamiento global intentará combinar los diagnósticos locales generados por el tercer nivel, en forma de un diagnóstico global de los problemas generales que presumiblemente están causando la aparición de las disfunciones locales detectadas.

Cada uno de estos pasos necesita de distintas técnicas para su resolución. En el resto de este trabajo se describirá la implementación del módulo de razonamiento global, es decir el nivel superior de la arquitectura, en forma de una red bayesiana. Para una descripción de los otros niveles se pueden consultar [Alamán 93], [Aguilar et al. 92] y [Aguirre et al. 92].

4. El nivel de razonamiento global

El nivel de razonamiento global elabora sus conclusiones a partir de los resultados obtenidos por el nivel de razonamiento local. El nivel de razonamiento local produce una serie de mensajes de diagnóstico de zonas muy localizadas de la planta. Dicho diagnóstico se realiza mediante un sistema basado en reglas, y las conclusiones, por la naturaleza local y concreta de

los problemas abordados, presentan un alto índice de robustez. Cuando el nivel de razonamiento local diagnostica la existencia de un problema, dicho problema está ocurriendo en la planta con casi absoluta certeza.

Un experto, viendo una lista de tales diagnósticos locales, puede muchas veces aventurar cuál puede ser la causa última que liga las deficiencias detectadas en diversos puntos de la planta. El nivel de seguridad con el que el experto realiza este diagnóstico global es mucho menor que en el caso de los diagnósticos locales. En esencia se trata de inferir un determinado problema global como la causa común de una serie de problemas locales, pero dejando siempre espacio para las excepciones, los casos no contemplados, etc.

Así planteado, el problema es similar al que se encuentra en medicina cuando se quiere encontrar cual es la enfermedad o conjunto de enfermedades que mejor explica los síntomas observados. Las relaciones causales presentes en ambos casos se prestan a ser representadas mediante un tipo especial de red bayesiana: las redes bipartitas.

Una red bayesiana es un grafo conexo dirigido, sin ciclos directos, y en el que los nodos representan conceptos elementales (predicados lógicos multivaluados), y los arcos que los unen relaciones de causalidad entre dichos predicados. Una red bayesiana bipartita es aquella en la que cada nodo o bien no tiene nodos padres (nodo causa, enfermedad) o bien no tiene nodos hijo (nodo efecto, síntoma). Para una discusión más detallada de los conceptos de red bayesiana y red bayesiana bipartita ver [Pearl 88].

El nivel de diagnóstico global puede ser entonces representado por una red como la que se muestra en la **figura 3**. Los nodos 'efecto' representan distintos problemas locales (detectados o no), y los nodos 'causa' distintas causas globales a las que se puede atribuir la aparición de los problemas locales. El cometido del nivel de razonamiento global es encontrar el conjunto de problemas globales que mejor explica el conjunto de problemas locales observados.

[Peng y Reggia 87] proponen un método probabilista para la resolución de este problema. El método se basa en una serie de asunciones de independencia, que sin embargo son bastante realistas para aplicaciones prácticas. Esencialmente se pide que los nodos causa (enfermedades) sean independientes entre sí, y que también lo sean las excepciones a que una enfermedad presente un determinado síntoma. El método se

basa en traducir el problema en términos de optimización, buscándose el mínimo mediante un modelo basado en competencia entre nodos. El método ofrece respuestas acertadas la mayor parte de las veces y con una eficiencia aceptable. El principal problema que aparece es que, algunas veces, el mínimo obtenido es un mínimo local, en contraposición al mínimo global buscado.

En [Pearl 88] se presenta un algoritmo alternativo para el cálculo de redes bayesianas, y por tanto para grafos bipartitos. Las asunciones del algoritmo son las mismas que las de Peng y Reggia, más la condición de que no existan ciclos (no dirigidos) en la red. En estas condiciones el algoritmo siempre obtiene la solución exacta en tiempo polinomial, mediante un mecanismo de intercambio de mensajes entre nodos. Naturalmente, el problema es que la mayor parte de los casos de interés real tienen ciclos, invalidando por tanto el algoritmo. Pearl propone diversos métodos para solucionar estos casos: simulación por Montecarlo, clustering y condicionamiento. Sin embargo, razones de índole práctica impiden que dichas soluciones puedan ser aplicadas en casos reales, debido a su ineficiencia.

En el módulo de razonamiento global que implementa el nivel superior de la arquitectura MIP se emplea el algoritmo de Pearl. El problema de los ciclos se aborda con una nueva aproximación: ignorar los ciclos, y estudiar en qué casos o bajo qué condiciones la solución (no exacta) obtenida se aproxima suficientemente a la solución exacta. El razonamiento que subyace a esta aproximación es que el efecto de un ciclo en el algoritmo de intercambio de mensajes es la introducción de un nivel de ruido sobre los mensajes que representarían la solución exacta. Este ruido podría ser aceptable si es suficientemente pequeño, especialmente si estamos interesados en resultados cualitativos (el diagnóstico más probable) y no en los valores de probabilidad concretos de cada hipótesis. De hecho, si dos soluciones difieren en su probabilidad en un valor muy pequeño - y por tanto el ruido introducido por los ciclos puede hacer confundir una solución por la otra - entonces la solución (errónea) obtenida resulta ser casi tan probable como la exacta, y por tanto el error cometido no es muy relevante.

Hay razones para creer que este es el caso de las redes bipartitas. La naturaleza de éstas imponen una estructura en la que los ciclos obligatoriamente alternan nodos causa y nodos consecuencia. Dado que en los nodos consecuencia la evidencia externa juega un papel dominante, la influencia negativa de los ciclos en la solución final está en cierta manera mediatizada.

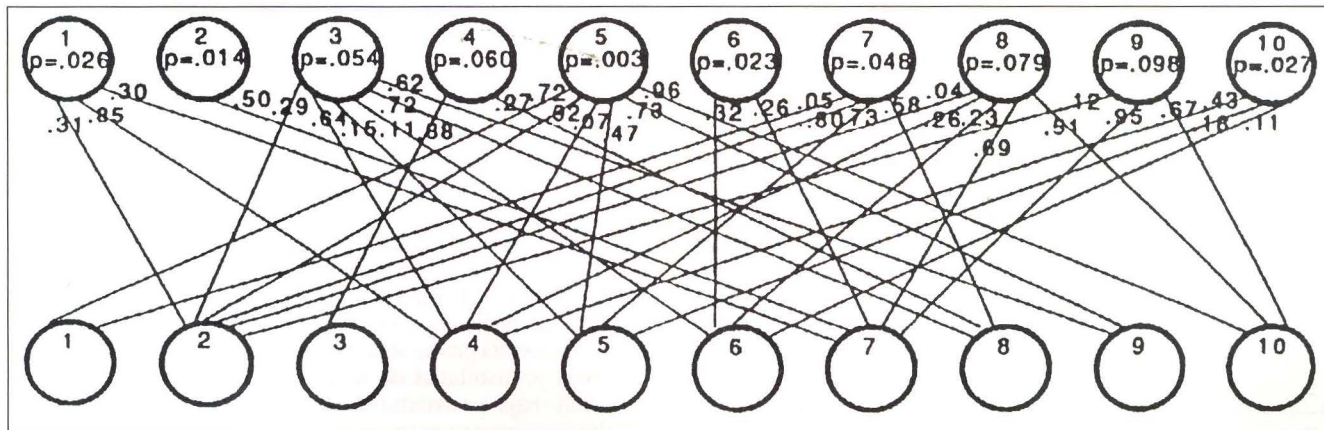


Figura 3: Ejemplo de red bipartita

Adicionalmente, en problemas de aplicación práctica habitualmente se busca el estado más probable del conjunto de causas que explica el conjunto de consecuencias. Esto significa emplear el modo de razonamiento de revisión de credibilidades, donde los sumatorios del algoritmo son sustituidos por operaciones de maximización (ver [Pearl 88]). Este modo de razonamiento puede comprobarse experimentalmente que es más robusto frente al 'ruido' introducido por los ciclos. Esto es debido a la naturaleza no lineal del algoritmo que en el estado final origina que la credibilidad del valor que pertenece a la solución más probable de la red sea muy superior, habitualmente, a la credibilidad de los otros valores.

La presencia de un pequeño ruido no cambiará la elección de dicho valor y por tanto no influirá en la decisión final. Dicho de otra forma, el algoritmo de revisión de credibilidades se comporta de una manera no lineal, ofreciendo bien una solución u otra completamente distinta. Esto origina que las soluciones correctas tengan bastante estabilidad frente a pequeñas variaciones (ruidos) en los parámetros, lo cual en la mayor parte de los casos anula por completo el efecto negativo de los ciclos.

Se han realizado tres experiencias realizadas para corroborar experimentalmente la viabilidad del algoritmo de relajación por intercambio de mensajes en redes bayesianas bipartitas. Las experiencias se han realizado contrastando los resultados de experimentos para redes bipartitas realizados por [Peng y Reggia 87] empleando un algoritmo conexionista basado en optimización local con los resultados obtenidos por el algoritmo de redes bayesianas sobre los mismos datos.

Con cada uno de ellos se realizó la siguiente experiencia: para cada posible combinación de síntomas (hay 1024 combinaciones) se inicializa la red bayesiana con la evidencia correspondiente. Seguidamente se ejecuta el algoritmo de relajación en el modo de revisión de credibilidad. Cuando la red alcanza una situación estable, se compara la solución calculada con la obtenida por Peng y Reggia así como con la solución exacta, calculada con un algoritmo de búsqueda exhaustiva. Para considerar el caso de que la red no converja, sino que se tenga una situación de oscilación (o en principio incluso de divergencia), la experiencia se realizó fijando un máximo de 200 iteraciones. Toda solución que no presentó convergencia dentro de este margen se consideró como no convergente.

En la **Tabla** se presentan los resultados obtenidos por Peng y Reggia y por el método bayesiano para cada uno de los tres experimentos. Cada columna resume los resultados obtenidos para las 1024 diferentes combinaciones de síntomas y para los tres experimentos, mostrándose el número de casos correctos resueltos y entre paréntesis el porcentaje que éstos representan sobre el total. En la primera columna se presentan los resultados del método conexionista de Peng y Reggia. La segunda columna corresponde a los resultados obtenidos por el algoritmo de redes bayesianas. La tercera columna presenta los resultados de Peng y Reggia cuando se permite intervención externa ante la sospecha de que se ha obtenido un mínimo local.

	P&R (conexionista)	INFER	P&R (no conexionista)
Ejemplo 1	927 (90.5%)	951 (92.9%)	1015 (99.1%)
Ejemplo 2	859 (83.9%)	909 (88.8%)	1019 (99.5%)
Ejemplo 3	894 (87.3%)	979 (95.6%)	1020 (99.6%)
Total	2680 (87.2%)	2839 (92.4%)	3054 (99.4%)

Los resultados obtenidos por el método bayesiano son bastante satisfactorios. Es justo realizar la comparación con los resultados del método conexionista de Peng y Reggia. En esta comparación el método bayesiano mejora al método de Peng y Reggia en todos los experimentos. Sin embargo, si se compara el método bayesiano con el método de Peng y Reggia con intervención externa, el segundo es sensiblemente más fiable. La razón de ello, naturalmente, reside en que el segundo método no es realmente un método conexionista. Si al método de relajación bayesiana se le añade la posibilidad de intervención externa (por ejemplo mediante un agente que escoja adecuadamente algunos nodos sobre los que aplicar el método de condicionamiento) sin duda podrían conseguirse resultados comparables o mejores.

Sin embargo, en principio, se prefiere no emplear esta aproximación: las propiedades conexionistas del sistema deben ser preservadas mientras sea posible. Hay que señalar que las dificultades encontradas por el método de Peng y Reggia son muy distintas en naturaleza de las encontradas por el método bayesiano. El método de Peng y Reggia siempre encuentra un mínimo: su problema es que éste puede ser un mínimo local, que no hay manera de diferenciar (de manera local) de un mínimo global. El método bayesiano en cambio encuentra problemas, la mayor parte de las veces, cuando el estado de la red no converge (habitualmente, cuando la red oscila entre varios estados metaestables). En principio el problema de decidir si una red bayesiana está en situación oscilatoria es mucho más sencillo que el decidir si un determinado mínimo es local o global.

De hecho, un interesante resultado experimental obtenido es que la mayor parte de los fallos del método bayesiano correspondieron a situaciones de no convergencia en vez de corresponder a situaciones de convergencia a una solución errónea. Esto corrobora la intuición inicial de que las soluciones obtenidas por revisión de credibilidades tienen que ser altamente robustas frente al ruido (y en caso de no serlo, quiere decir que se está muy cerca del punto de frontera de dos estados distintos, lo cual implica efectivamente que el ruido introducido por los ciclos pueda causar una oscilación entre ambos). De un total de 233 casos no solucionados, 217 correspondieron a estados no convergentes de la red, mientras que sólo en 16 casos se encontró convergencia a una solución errónea, y en estos últimos la solución obtenida resultó ser una solución admisible, aunque no óptima, del problema.

Si se aceptan los resultados de estos experimentos como representativos, el método de relajación de redes bayesianas parece una aproximación muy fiable al problema de diagnóstico en grafos bipartitos. Por ejemplo, en los problemas considerados en los anteriores experimentos cuando se encontró un estado convergente el porcentaje de aciertos fue muy alto (99.4% de aciertos, esto es 2839 soluciones correctas entre 2855 casos en los que la red presentó convergencia). Cuando el algoritmo no convergió no se encontró la solución, pero al menos se sabía que éste era el caso, lo cual es fundamental para aplicaciones prácticas. Es más, esta situación se da con una baja frecuencia (en nuestros experimentos solo un 8.5% de las veces).

En resumen, el método bayesiano permite resolver aproximadamente el problema de diagnóstico en grafos bipartitos con una alta probabilidad de encontrar la solución correcta, una baja probabilidad de no encontrar ninguna solución, y una muy baja probabilidad de encontrar una solución que resulte ser errónea (en los experimentos, del orden de 0.0005).

Finalmente, es interesante comentar algunos resultados relacionados con la convergencia de las soluciones. En los experimentos realizados el 58.4% de los casos empleó entre 4 y 9 ciclos para converger, el 29.9% de los casos empleó entre 10 y 19 ciclos y sólo 3.2% más de 20 ciclos. En posteriores experimentos de control no se encontró ningún caso que necesitara más de 71 ciclos para converger (es decir, los casos no convergentes mantienen su carácter no convergente a pesar de que se aumente considerablemente el número de ciclos permitidos).

Tres tipos de problemas de convergencia se encontraron en los experimentos: soluciones oscilatorias, convergencia a soluciones no óptimas, y convergencia dependiente del orden de activación de los nodos en el algoritmo.

Los problemas de oscilación parecen ser los más frecuentes. La oscilación suele tomar la forma de dos estados metaestables de la red, muchas veces presentando algún tipo de simetría (por ejemplo, presentando algunos parámetros con igual valor y otros parámetros con valores intercambiados). Los estados metaestables se alternan indefinidamente en el tiempo. No existe razón por la que en principio no se puedan obtener oscilaciones entre tres o más estados metaestables, pero en los experimentos realizados no se ha encontrado tal situación.

El segundo problema aparece cuando la red bayesiana converge a una solución errónea. Esta es una situación realmente grave, pues no hay manera de distinguirla de la obtención de una solución correcta. Afortunadamente, según se comprueba de modo experimental, este caso es muy poco frecuente.

El tercer problema tiene que ver con la sensibilidad del algoritmo al orden en que son activados los nodos durante la relajación. Efectivamente se ha encontrado que en algunos casos el orden de activación de los nodos ha tenido influencia sobre la convergencia o no de la red. Siempre que esto ha ocurrido, la transición ha sido entre una solución exacta y un estado de no convergencia. Es un resultado interesante que distintos ordenes en la activación de los nodos conduzcan a que la no convergencia se presente en distintos supuestos de ordenación. Esta dependencia, que en principio es una propiedad negativa del algoritmo de relajación, sin embargo puede ser empleada en nuestro beneficio: si para un determinado caso el algoritmo no converge, se puede intentar emplear un orden distinto de activación de los nodos.

5. Implementación y resultados

El éxito alcanzado por el algoritmo de relajación bayesiana para redes bipartitas en los experimentos de control, sugirió la posibilidad de su empleo en el módulo de razonamiento global de la arquitectura MIP. Dicho razonamiento global puede ser reformulado en el lenguaje de las redes bayesianas como un problema de revisión de credibilidad en una red bipartita.

Efectivamente, se dispone con seguridad de información acerca de la presencia o no de una serie de síntomas relativos al estado de un proceso continuo determinado ('diagnósticos locales'). El módulo de razonamiento global tiene que deducir el conjunto de causas globales que mejor explican estos diagnósticos locales (síntomas) observados.

Dos problemas aparecen al realizar esta tarea. Primero hay que resolver la coincidencia temporal de los distintos síntomas. Segundo hay que realizar un razonamiento diagnóstico para encontrar el conjunto de causas globales que mejor explican los distintos problemas locales observados.

El primer aspecto se resuelve dando a cada problema local (diagnóstico del módulo de razonamiento local) un tiempo de vida limitado, relacionado con el concepto de perdurabilidad del correspondiente problema. De esta manera, cada diagnóstico local irá teniendo una influencia menor según transcurra el tiempo. El problema del razonamiento diagnóstico se resuelve mediante una red bayesiana bipartita a partir de los valores de evidencia que aportan los diagnósticos locales presentes en la pizarra. El resultado del razonamiento se envía a la pizarra, de donde lo obtendrá el interfaz de usuario para su presentación al operador.

El método descrito ha sido implementado como el nivel superior de una aplicación industrial real. El número de posibles diagnósticos locales en dicha aplicación es 70, causados por 19 posibles problemas globales. Un total de 497 variables de la planta son consideradas para la realización del diagnóstico, incluyendo 161 variables calculadas.

Para la obtención del conocimiento de este nivel fueron discutidos con los expertos del proceso los problemas generales de la planta más representativos, clasificados dentro cuatro clases de problemas: desajustes en la reacción, fallos de instrumentación, problemas de refrigeración y problemas de flujo del aire. Sólo fueron analizados los problemas globales que, debido a su mayor frecuencia de aparición o a su repercusión económica, se juzgaron de mayor impacto para la planta de ACN.

Para cada problema global se realizó un estudio de los efectos posibles dentro de cada uno de los subsistemas de la planta, tanto desde el punto de vista de los modelos existentes, como desde el punto de vista de la experiencia práctica del ingeniero de procesos. Dichas repercusiones se expresaron en forma de los diagnósticos locales que el módulo de razonamiento local produce en cada caso. De esta manera se estimó para cada problema global el conjunto de diagnósticos locales que era previsible que fueran generados, junto con un coeficiente de probabilidad para tal relación causal. Dada la carencia de datos históricos suficientes para calcular estas probabilidades con técnicas estadísticas, se recurrió a la experiencia del ingeniero de procesos para su estimación. El conjunto de problemas globales estudiados y sus diagnósticos locales correspondientes conformaron una red bayesiana con más de 63 nodos y del orden de 100 relaciones causales.

La técnica de razonamiento bayesiano demostró ser muy adecuada para la implementación de este nivel del conocimiento dentro de la arquitectura propuesta. Desde el punto de vista de la adquisición de conocimiento, la estructuración en forma de problemas globales causando efectos locales en subsistemas de la planta resulta natural para el ingeniero de procesos, que puede identificar sin esfuerzo tales relaciones causales. La dificultad en la cuantificación de cada relación causal está suavizada por la introducción de una escala lingüística. La falta de precisión resultante no parece ser demasiado significativa, dado el carácter aproximado del posterior razonamiento, así como la robustez de la solución frente a pequeñas modificaciones en los parámetros del sistema. El razonamiento resultante mostró todas las cualidades que caracterizan el mecanismo de relajación de redes bayesianas.

Los resultados obtenidos son aproximados (ya que hay ciclos no dirigidos en la red bayesiana), pero muestran un grado de robustez adecuado. El carácter no excesivamente conexo de la red, que probablemente es característico de muchos problemas en control de procesos, junto con las ventajas inherentes al razonamiento bayesiano, permiten la preservación de algunas

propiedades esenciales del razonamiento del experto.

Una de las propiedades más útiles entre las exhibidas por el módulo de razonamiento global es la capacidad de razonamiento por defecto (explaining away). Por ejemplo, considérese el caso de los problemas globales 3.1 y 3.2, mostrados en la **figura 4**. Si se generan únicamente los diagnósticos locales identificados por los números 61 y 62, el módulo de razonamiento global inferirá que la causa más probable de éstos es el problema global 3.2, ya que la no aparición de los diagnósticos 58 y 59 restarán credibilidad al problema 3.1, mientras que el problema 3.2 por sí solo explica completamente las evidencias conocidas. Si ahora el módulo de razonamiento local genera los diagnósticos 58 y 59, el módulo de razonamiento global 'cambiará de idea', y juzgará que el problema global existente es el 3.1. Esto es así porque la evidencia existente refuerza la creencia en el problema 3.1, y éste explica (explains away) toda la evidencia por sí solo, causando una pérdida de credibilidad en el problema 3.2. Este mecanismo de razonamiento por defecto y explaining away se aplica igualmente a casos más complejos, donde la justificación entre causas alternativas se amplía a varios niveles y a múltiples causas simultáneas.

Otra característica que presenta el mecanismo de razonamiento bayesiano es su interruptibilidad. Al ser un algoritmo de convergencia, éste puede ser interrumpido en cualquier momento. Naturalmente la calidad de las estimaciones probabi-listas será inferior si no se permite un tiempo suficiente de convergencia, pero éste es el precio de querer una respuesta extremadamente rápida. El algoritmo asegura, en todo caso, la mejora incremental de la solución en el tiempo, propiedad ésta de la mayor importancia para algoritmos interrumpibles.

El método ha demostrado ser muy robusto, habiendo sido validados positivamente los resultados del módulo de razonamiento global contra un emulador empleado para la verificación y validación del sistema industrial. El sistema MIP está en producción en la planta desde marzo de 1991, habiéndose reportado una mejora en la producción de un 4%, aparte de otros beneficios relacionados con mejor uso del equipamiento, mejora de la eficiencia de los operadores, y evitamiento de situaciones de emergencia.

Agradecimientos: En primer lugar quiero agradecer a Erik de Pablo, coordinador y principal responsable del proyecto MIP, todas las discusiones e ideas que han dado forma a este trabajo. Rafael Muñoz y Sesh Murthy participaron activamente en las sesiones de adquisición del conocimiento para el módulo de razonamiento global. Otros participantes en el proyecto MIP fueron Soledad Romero, Carlos Aguirre, Pedro Serrahima, Vicente López, José Dorron-soro and Juan Antonio Aguilar. Agradezco a todos ellos su trabajo y contribuciones durante el desarrollo del proyecto.

Referencias

Aguilar, J.A., E. de Pablo, X. Alamán, (1992). "A Fuzzy Logic Approach for Sensor Validation in Real-Time Expert Systems", Intl. Conf.on Inf. Proc. and Management of Uncertainty in K.B. Systems, Palma de Mallorca, Spain, July 1992.
Aguirre, C., E. de Pablo, J.L. Zaccagnini, X. Alamán, (1992). "The User Interface in Expert Systems for Real-Time Process Control: The MIP system experience", 12th. IFIP World Computer Congress, Vol.3, pp. 266-272, Madrid, 1992.
Alamán, X., (1993). "Una arquitectura de integración para sistemas inteligentes de control", tesis doctoral presentada en el Dº de Informática y Automática de la U.C.M., abril 1993.

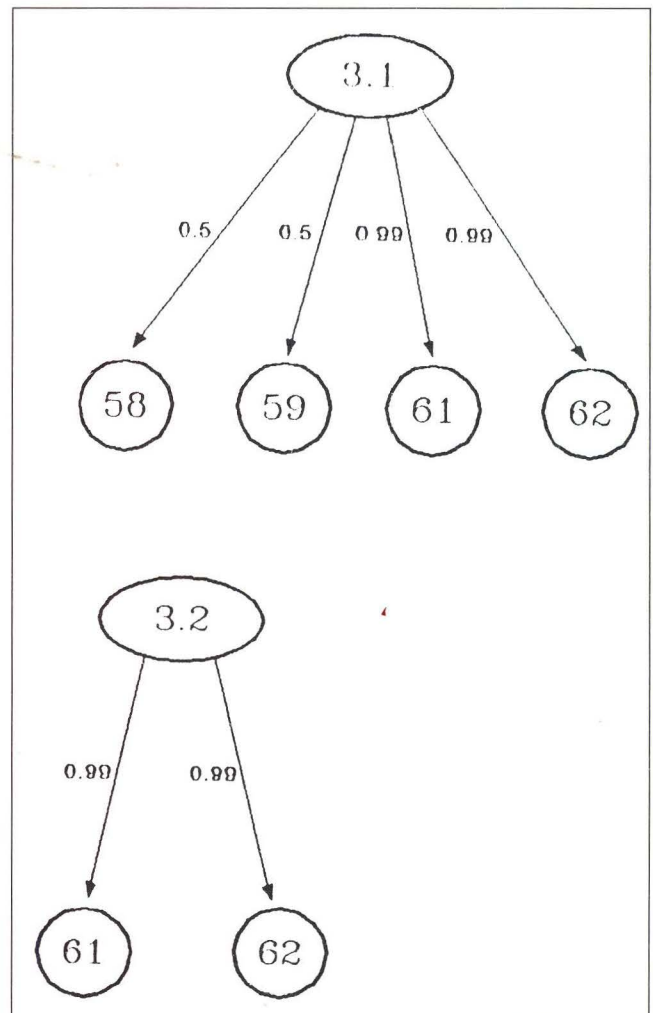


Figura 4: Un caso real

Alamán, X., S. Romero, C. Aguirre, P. Serrahima, R. Muñoz, V. López, J. Dorron-soro, E. de Pablo, (1991). "Monitorización Inteligente de Procesos: El Proyecto MIP", Actas de la IV Reunión Técnica de la Asoc. Española para la I.A.(AEPIA-91), pp. 423-432, Madrid, Octubre 1991.
Alamán, X., S. Romero, C. Aguirre, P. Serrahima, R. Muñoz, V. López, J. Dorron-soro, E. de Pablo, (1992). "Knowledge-Based Systems for Real-Time Process Control: The MIP Project, 1992 IFIP/IFAC/IMACS", Intl. Symposium on A.I. in Real-Time Control, Delft, Netherlands, June 1992.
Chang, G., J.D. Birdwell, (1990). "An Expert System Supervisor in Real-Time Control", 29th Conference on Decision and Control, Honolulu, Hawaii, Diciembre 1990.
Efstathiou, J., (1985). "Rule-based Process Control, Expert Systems and Optimisation in Process Control", A.Mamdani, J.Efstathiou(Eds), Technical Press.
Engelmore, R., T. Morgan, (1988). "Blackboard Systems", Addison-Wesley, California.
Erman, L.D., F. Hayes-Roth, V.R. Lesser, D.R Reddy, (1980). "The Hearsay-II speech understanding system: integrating knowledge to resolve uncertainty", ACM Computing Surveys, Vol 12, No. 2, pp. 213-253.
Pearl, J., (1988). "Probabilistic Reasoning in Intelligent Systems", Morgan Kauffman Pub., S. Mateo.
Peng, Y., and J.A. Reggia (1987). "A Probabilistic Causal Model for Diagnostic Problem Solving", IEEE Trans. on Systems, Man and Cybernetics, Vol. SMC-17.