

# Estudio de sistemas semicualitativos con restricciones. Aplicación a un modelo logístico con retraso

J.A. Ortega, R.M. Gasca, M. Toro, J. Torres y M.G. Romano  
Departamento de Lenguajes y Sistemas Informáticos.  
Universidad de Sevilla  
Avda. Reina Mercedes s/n. Sevilla (Spain)  
{ortega, gasca,mtoro}@lsi.us.es

## Resumen

*En este trabajo se propone una nueva metodología para automatizar el análisis y el estudio de modelos con restricciones que integran conocimiento cuantitativo y cualitativo construidos para sistemas que evolucionan en el tiempo.*

*Se hace un estudio detallado de la aplicación de la metodología a un modelo logístico al que se le incorpora un retraso en el bucle de realimentación del sistema.*

**Palabras clave:** *Simulación y razonamiento semicualitativo, Representación del conocimiento, Modelo logístico con retraso*

## 1 Introducción.

Los modelos de los sistemas que evolucionan en el tiempo y que se estudian en ciencia e ingeniería están normalmente compuestos de conocimiento cuantitativo, cualitativo, y semicualitativo. Todo este conocimiento debe ser tenido en cuenta cuando estos modelos son estudiados. En la bibliografía se han considerado diferentes niveles de abstracción numérica, desde una descripción: puramente cualitativa [Kuipers94], semicualitativa [Kay96], [Ortega *et al.* 98], [Ortega00] numérica intervalar [Vescovi *et al.* 95] y cuantitativa.

Se han desarrollado diferentes aproximaciones en la bibliografía cuando el conocimiento cualitativo se tiene en cuenta en los modelos: transformación de relaciones no lineales a lineales a trozo, método de MonteCarlo, programación lógica con restricciones, distribuciones de probabilidad, relaciones causales, conjuntos borrosos y combinación de todos los niveles de abstracción cualitativa cuantitativa. Una descripción detallada de todas estas técnicas se encuentra en [Ortega00].

En este trabajo se pretende estudiar sistemas con restricciones que evolucionan en el tiempo y que se componen de conocimiento cualitativo y cuantitativo. Los modelos que se construyen para estos sistemas se pueden formular con conocimiento cualitativo, operadores aritméticos y relaciones, funciones predefinidas ( $\ln$ ,  $x^p$ ,  $s^n$ , ...), literales numéricos e intervalos reales.

La metodología propuesta trata formalmente un modelo semicualitativo en una familia de modelos cuantitativos. La simulación de cada modelo cuantitativo genera una trayectoria en el espacio de estados.

Estos comportamientos cuantitativos se almacenan en una base de datos. Se propone un lenguaje de consulta/clasificación sobre esta base de datos. Las consultas permiten conocer las propiedades cualitativas de las trayectorias de la base de datos. El lenguaje también permite clasificar los diferentes comportamientos cualitativos del sistema. Este comportamiento se expresa por medio de un conjunto de reglas cualitativas jerárquicas. Las secciones siguientes describen la metodología en detalle.

Un estudio teórico sobre la validez de las conclusiones obtenidas del modelo y su extrapolación al sistema real se encuentra en [Ortega00]. Finalmente se aplica la metodología a un modelo de crecimiento logístico con un retraso.

## 2 Metodología.

La metodología propuesta permite estudiar tanto el régimen estacionario como el régimen transitorio de los sistemas que evolucionan en el tiempo. La principal idea de la metodología es: "cuando un modelo semicualitativo se transforma en una familia de modelos cuantitativos, cada uno de estos modelos tiene un comportamiento cuantitativo diferente, si bien entre si pueden responder a comportamientos cualitativos similares".

La metodología permite razonar sobre los modelos semicualitativos con restricciones. Para conseguirlo, combina de cierta manera y en un cierto orden: reglas de transformación, técnicas estocásticas, simulación cuantitativa, generación de bases de datos, lenguajes de consulta/clasificación sobre bases de datos, minería de datos, técnicas de clasificación y agrupamiento o *clustering*, y algoritmos para obtención de patrones de comportamiento. El detalle sobre la manera y el orden de aplicar estos pasos se describe conforme se avanza en los capítulos de esta memoria.

La metodología se muestra en la Fig. 1.

Sea un modelo semicualitativo  $S$  obtenido de un sistema dinámico con conocimiento cualitativo, formulado como sigue

$$\dot{\Phi}(\dot{x}, x, q, t), \quad x(t_0) = x_0, \quad \Phi_0(q, x_0) \quad (1)$$

siendo  $x \in \mathbb{R}^n$  el conjunto de variables de estado del sistema,  $q$  los parámetros,  $t$  el tiempo,  $\dot{x}$  la variación de las variables de estado con el tiempo,  $\Phi_0$  las restricciones con las condiciones iniciales del sistema, y  $\Phi$  las restricciones que definen el sistema y que dependen de  $\dot{x}, x, q, t$ .

Cuando la metodología se aplica, las ecuaciones del sistema dinámico (1) se transforman en un conjunto de restricciones entre las variables, parámetros e intervalos. En este artículo, estamos interesados en aquellos sistemas que pueden expresarse como (2) después de que las reglas de transformación se hayan aplicado

$$\dot{x} = f(x, p, t), \quad x(t_0) = x_0, \quad p \in I_p, \quad x_0 \in I_0 \quad (2)$$

donde  $p$  incluye los parámetros del sistema y los nuevos parámetros obtenidos por medio de las reglas de la transformación que posteriormente explicaremos en este trabajo,  $f$  es una función obtenida aplicando las reglas de transformación, e  $I_p, I_0$  son intervalos reales. La ecuación (2) es una familia de sistemas dinámicos dependientes de  $p$  y  $x_0$ .

Aplicando un conjunto de reglas de transformación descritas por la metodología, el modelo  $S$  se transforma en una familia de modelos cuantitativos  $F$ . Cada modelo cuantitativo  $M \in F$  se

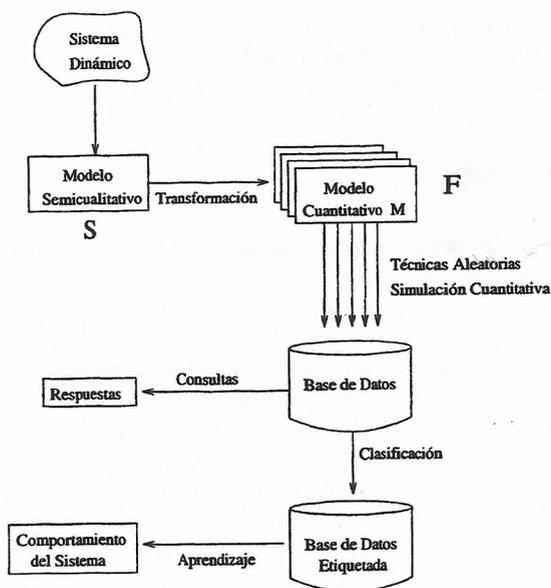


Figura 1: Metodología propuesta

obtiene aplicando técnicas estocásticas. La simulación cuantitativa de  $M$  origina una trayectoria cuantitativa. Cada una de ellas se forma con los valores que toman las variables de estado desde su valor inicial hasta su valor final y los valores de los parámetros del sistema. Por lo tanto, cada trayectoria contiene para las variables de estado, sus valores en régimen transitorio y en estacionario. Se obtiene una base de datos de trayectorias cuantitativas  $T$  con todos estos comportamientos cuantitativos. Se propone un lenguaje para realizar preguntas sobre las propiedades cualitativas del conjunto de trayectorias incluidos en la base de datos. Estas trayectorias también pueden ser clasificadas según algún criterio, obteniéndose entonces una base de datos etiquetada. Los diferentes patrones cualitativos de comportamiento del sistema se pueden obtener automáticamente a partir de la base de datos etiquetada aplicando algoritmos de aprendizaje basados en algoritmos genéticos. En las secciones siguientes, vamos a describir los pasos de la metodología en detalle. En la siguiente sección se muestran el tipo de conocimiento cualitativo y sus técnicas de la transformación. A continuación, se explica el algoritmo para la generación de la base de datos, y finalmente, la sintaxis abstracta y la semántica del lenguaje de consulta/clasificación se define.

### 3 Conocimiento cualitativo.

La representación del conocimiento cualitativo se realiza mediante operadores que tienen asociados intervalos reales. La ventaja principal de esta representación es que la integración del conocimiento cualitativo y cuantitativo se hace de una manera simple, facilitando además la incorporación de conocimiento de los expertos en la definición del rango de las variables y los parámetros cualitativos del sistema [Gasca98].

El conocimiento cualitativo se puede componer de operadores cualitativos, etiquetas cualitativas,

funciones de banda y funciones continuas cualitativas. Cada uno de ellos y sus técnicas de la transformación se detallan a continuación.

### 3.1 Operadores Cualitativos

La representación de parámetros y condiciones iniciales cualitativas se representa mediante los *operadores cualitativos*. Cada operador *op* cualitativo se define mediante un intervalo  $I_{op}$  proporcionado por los expertos. Se clasifican en unarios  $U$  y binarios  $B$ .

**Operadores unarios.** Cada magnitud cualitativa del sistema tiene sus propios operadores unarios. Sea  $U_x$  el conjunto de operadores unarios para una magnitud  $x$ , es decir  $U_x = \{VN_x, MN_x, LN_x, AP0_x, LP_x, MP_x, VP_x\}$ . Representan para  $x$  las etiquetas cualitativas *muy negativo, moderadamente negativo, ligeramente negativo, aproximadamente cero, ligeramente positivo, moderadamente positivo y muy positivo* respectivamente. La regla de la transformación cada operador unario es la siguiente

$$op_u(e) \equiv \left\{ e - r = 0 \quad r \in I_u \right. \quad (3)$$

siendo  $r$  una nueva variable auxiliar e  $I_u$  el intervalo asociado al operador  $op_u$  que se determina de acuerdo con [Travé-Massuyès *et al.* 97].

**Operadores binarios.** Sean  $e_1, e_2$  dos expresiones aritméticas, y sea  $op_b$  un operador binario. La expresión  $op_b(e_1, e_2)$  denota una relación cualitativa entre  $e_1$  y  $e_2$ . Los operadores binarios se clasifican en

- Operadores relacionados con la diferencia  $\geq, =, \leq$ , a los que se les aplican las siguientes reglas de transformación

$e_1 = e_2 \equiv e_1 - e_2 = 0$	$e_1 \leq e_2 \equiv \left\{ \begin{array}{l} e_1 - e_2 - r = 0 \\ r \in [-\infty, 0] \end{array} \right.$	$e_1 \geq e_2 \equiv \left\{ \begin{array}{l} e_1 - e_2 - r = 0 \\ r \in [0, +\infty] \end{array} \right.$
----------------------------------	--	--

- Operadores relacionados con el cociente  $\{\ll, - <, \sim, \approx, \gg, Vo, Ne, \dots\}$ . La regla de transformación es la siguiente

$$op_b(e_1, e_2) \equiv \left\{ \begin{array}{l} e_1 - e_2 * r = 0 \\ r \in I_b \end{array} \right. \quad (4)$$

siendo  $r$  una nueva variable e  $I_b$  el intervalo asociado a operador  $op_b$  de acuerdo con [Travé-Massuyès *et al.* 97]

### 3.2 Funciones de banda

Una función de banda  $y = g(x)$  (figura 2.a) representa la familia de funciones incluidas entre dos funciones reales, una superior  $\bar{g} : \mathbb{R} \rightarrow \mathbb{R}$  y otra inferior  $\underline{g} : \mathbb{R} \rightarrow \mathbb{R}$ . Se denota por medio de

$$\langle \underline{g}(x), \bar{g}(x), I \rangle, \quad \forall x \in I : \underline{g}(x) \leq \bar{g}(x) \quad (5)$$

donde  $I$  es el rango de definición de  $g$ , y  $x$  es la variable independiente. La regla de la transformación que se aplica a (5) es

$$g(x) = \alpha \underline{g}(x) + (1 - \alpha) \bar{g}(x) \quad \text{con } \alpha \in [0, 1] \quad (6)$$

donde  $\alpha$  es una nueva variable.

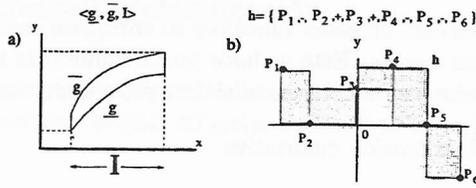


Figura 2: Funciones cualitativas

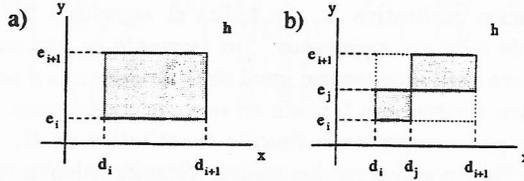


Figura 3: Interpretación cualitativa de una función

### 3.3 Funciones continuas cualitativas

Una función continua cualitativa  $y = h(x)$  (figura 2.b) representa una restricción entre los valores de  $x, y$  según las propiedades de  $h$ . Se denota por

$$y = h(x), \quad h \equiv \{P_1, s_1, P_2, \dots, s_{k-1}, P_k\} \quad (7)$$

siendo  $P_i$  puntos de la función, estando cada uno definido por el par de marcas cualitativas  $(d_i, e_i)$ , donde  $d_i$  es la marca asociada a la  $x$  y  $e_i$  a  $y$ . Estos puntos están separados por el signo  $s_i$  de la derivada en el intervalo entre dos puntos consecutivos. Una función cualitativa monótona es un caso particular de estas funciones donde el signo es siempre el mismo  $s_1 = \dots = s_{k-1}$ .

La interpretación cualitativa (figura 3.a) para cada  $P_i = (d_i, e_i)$  de  $y = h(x)$  es

$$y - h(x) = 0 \equiv \begin{cases} \text{si } x = d_i \Rightarrow y = e_i \\ \text{si } d_i < x < d_{i+1} \Rightarrow \begin{cases} s_i = + \Rightarrow e_i < y < e_{i+1} \\ s_i = - \Rightarrow e_i > y > e_{i+1} \\ s_i = 0 \Rightarrow y = e_i \end{cases} \end{cases} \quad (8)$$

Las reglas de transformación de una función continua cualitativa se aplican en tres pasos:

a) **Normalización** La definición de la función se *completa* y *homogeneiza* usando las siguientes propiedades de continuidad:

1. una función que cambia de signo entre dos puntos consecutivos pasa a través de un punto cuyo valor en la función es 0
2. una función cuya derivada cambia de signo entre dos puntos consecutivos, pasa por un punto cuya derivada es cero.

La definición de cualquier función (7) se completa siempre con los puntos extremos  $(-\infty, +\infty)$ , los puntos de cortes con los ejes, y con los puntos donde cambia la derivada (un máximo o un mínimo de  $h$ ).

b) **Extensión** La definición de estas funciones se enriquece mediante un proceso automático de incorporación de nuevas marcas. Esto se hace para disminuir la incertidumbre en la definición de una función, ya que como puede observarse el área del rectángulo se reduce (figura 3.b).

c) **Transformación** Una función cualitativa

$$y = h(x), \quad h \equiv \{P_1, s_1, P_2, \dots, s_{k-1}, P_k\} \text{ con } P_i = (d_i, e_i) \quad (9)$$

se transforma en un conjunto de funciones cuantitativas  $H$ . Cada función de  $H$  cumple las restricciones de la función cualitativa  $h$ . Se aplica el algoritmo **Seleccionar H** para obtener  $H$ . Este algoritmo divide  $h$  en sus segmentos. Un *segmento* es una sucesión de puntos consecutivos  $\{P_m, \dots, P_n\}$  de manera que todos tengan igual signo de derivada y no contengan un punto que corte un eje. Los segmentos dividen una función en regiones monótonas. El algoritmo propuesto aplica técnicas estocásticas para escoger cada función cuantitativa de  $H$ . Estas técnicas son similares al método de MonteCarlo, sin embargo, los valores obtenidos deben satisfacer las restricciones de  $h$ . La heurística seguida aplica una distribución uniforme aleatoria para obtener el valores para cada marca. Los valores obtenidos deben verificar las mismas relaciones de orden que existían entre sus respectivas marcas.

## 4 Base de datos.

Una familia  $F$  de modelos cuantitativos se obtiene tras aplicar las reglas de transformación a  $S$ . Esta familia depende de un conjunto de parámetros intervalares  $p$  y de las funciones  $H$ . Cada modelo particular  $M$  de  $F$  se selecciona por medio de técnicas estocásticas, y se simula de manera cuantitativa. Esta simulación genera una trayectoria  $r$  que se almacena en la base de datos  $T$ .

Los siguientes algoritmos se aplican para obtener  $T$ .

### SeleccionarModelo( $F$ )

para cada parametro o variable intervalar de  $F$   
 seleccionar un valor para el en su intervalo  
 para cada funcion  $h$  de  $F$   
 seleccionar una funcion cuantitativa  $H$

### Generacion de base de datos $T$

$T := \{\}$   
 para  $i=1$  a  $N$   
 $M := \text{SeleccionarModelo}(F)$   
 $r := \text{SimulacionCuantitativa}(M)$   
 $T := T \cup r$

siendo  $N$  el número de simulaciones a realizar. Este número se define de acuerdo con la sección 6, siendo por tanto  $N$  el número de trayectorias de  $T$ .

## 5 Lenguaje de consulta/clasificación.

En esta sección, se propone un lenguaje para realizar consultas sobre la base de datos de trayectorias. Este lenguaje también sirve para asignar etiquetas cualitativas a las trayectorias.

### 5.1 Sintaxis abstracta

$Q:$	$\forall r \in T \bullet [r, P]$	$P:$	$P_b$	$P_b:$	$P_d$	$P_d:$	$EQ$	$F:$	$F_b$	$F_b:$	$e_b$
	$\exists r \in T \bullet [r, P]$		$P \wedge P$		$f(L(F))$		$CL$		$F \& F$		$e \in I$
	$\mathcal{N}r \in T \bullet [r, P]$		$P \vee P$		$\forall t: F \bullet F$				$F   F$		$u(e)$
	$[r, p]$		$\neg P$		$\exists t: F \bullet F$				$!F$		$b(e, e)$

Sea  $T$  el conjunto de todas las trayectorias  $r$  almacenadas en la base de datos. Una consulta  $Q$  es: un cuantificador  $\forall, \exists, \mathcal{N}$  aplicado sobre  $T$ , o una consulta básica  $[r, P]$  que evalúa a *true* cuando la trayectoria  $r$  verifica la propiedad  $P$ .

Una propiedad  $P$  se puede formular mediante la composición de otras propiedades utilizando operadores lógicos  $\wedge, \vee, \neg$ . El resultado es la aplicación de los operadores entre las propiedades parciales.

Una propiedad básica  $P_b$  puede ser: una propiedad predefinida  $P_d$ , una función booleana  $f$  aplicado sobre una lista  $L$  de puntos o intervalos que verifican una fórmula  $F$ , o un cuantificador  $\forall, \exists$  aplicado a los valores de una trayectoria particular durante un tiempo  $t$ . Este tiempo puede ser: un instante, un rango de tiempo, un operador temporal predefinido, o la lista de tiempos donde se verifica una fórmula  $F$ .

Una propiedad predefinida  $P_d$  es aquella cuya formulación es automática. Corresponden normalmente con consultas que se realizan sobre los sistemas dinámicos. Hay dos predefinidas:  $EQ$  que se verifica cuando cuando la trayectoria acaba en un equilibrio estable; y  $CL$  cuando acaba en un ciclo límite.

Una fórmula  $F$  puede componerse de otras fórmulas combinadas por medio de operadores booleanos  $\&, |, !$ . Una fórmula básica  $F_b$  puede ser una expresión booleana  $e_b$ , o una expresión numérica  $e$  perteneciente a un intervalo, o un operador cualitativo unario  $u$  o binario  $b$ .

#### Clasificación

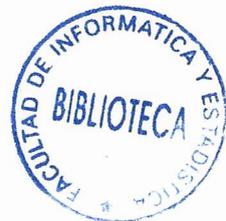
Una regla de la clasificación se formula como un conjunto de consultas básicas con etiquetas, y posiblemente otras expresiones

$$[r, P_A] \Rightarrow A, e_{n1}, \dots \quad [r, P_B] \Rightarrow B, e_{n2}, \dots \quad \dots$$

Una trayectoria  $r$  se clasifica con la etiqueta  $\delta$  si verifica la propiedad  $P_\delta$ .

### 5.2 Semántica

La semántica de cada instrucción de este lenguaje se transforma en una consulta sobre la base de datos. La consulta  $[r, P]$  es *true* si la trayectoria  $r$  verifica la propiedad  $P$ . La semántica de una consulta con un cuantificador depende de este cuantificador. Si es  $\forall$ , evalúa a *true* cuando todas



las trayectorias  $r \in T$  verifican  $P$ . Si es  $\exists$  devuelve *true* cuando hay al menos una trayectoria  $r \in T$  que verifica  $P$ . Si es  $\mathcal{N}$  entonces se devuelve el número de trayectorias de  $T$  que cumplen  $P$ .

Sea  $\forall t : F_1 \bullet F_2$  una propiedad básica la cual será *true* si durante el tiempo que  $F_1$  se satisface, todos los valores de  $r$  verifican  $F_2$ . Si el cuantificador es  $\exists$ , entonces es *true* cuando al menos un valor de  $r$  que satisface  $F_1$ , también satisface  $F_2$ . Para evaluar una fórmula  $F$ , es necesario sustituir sus variables para sus valores. Estos valores se obtienen de  $T$ .

Sea  $[elr, P_A] \Rightarrow A, e_{A1}$  una regla de la clasificación. Una trayectoria  $r \in T$  se clasifica con la etiqueta  $A$  si verifica propiedad  $P_A$ . El resultado de evaluar  $e_{A1}$  para esta trayectoria también se guarda en la base de datos.

## 6 Modelo logístico con retraso.

En el mundo real es muy común encontrar procesos de crecimiento donde una fase inicial de crecimiento exponencial es seguida por otra fase de acercamiento asintótico a un valor de saturación. Este comportamiento se produce en aquellos sistemas en los que su expansión exponencial se trunca por la limitación de los recursos necesarios para el crecimiento. A estos procesos se le dan los nombres genéricos: logístico, sigmoidal, o *s-shaped*.

Una característica común a todos estos sistemas es que la población se encuentra inicialmente en un entorno propicio que posibilita la etapa de crecimiento exponencial, pero el propio proceso de multiplicación de la población origina posteriormente la reducción del proceso de crecimiento.

El sociólogo P.F. Verhulst fue el primero en modelar estos sistemas basándose en el crecimiento de la población humana. A partir de entonces estos modelos se han estudiado profusamente en la literatura, descubriéndose que abundan en procesos naturales, sociales y socio-tecnológicos. Por citar algunos, aparecen en: la evolución de las bacterias, la extracción de mineral, el crecimiento de la población mundial, desarrollos económicos, las curvas de aprendizaje, determinados fenómenos dentro de una población, como rumores o epidemias, nuevos productos que se introducen en el mercado, etc.

Relacionando estos modelos logísticos con la dinámica de sistemas, se concluye que estos procesos son el resultado de la interacción entre un bucle de realimentación positiva que es dominante en la etapa inicial y por lo tanto del crecimiento inicial, y de un bucle de realimentación negativa que se encarga de la estabilización final y que es dominante en la etapa final del comportamiento [Richardson95].

El cambio en el dominio del bucle de realimentación está relacionado con las no linealidades de la estructura del modelo. Estas no linealidades, bajo ciertas circunstancias motivadas por los valores de los parámetros pueden causar un patrón de comportamiento bimodal en los que existen dos atractores en régimen estacionario del sistema (figura 4):  $A$  crecimiento logístico normal, y  $O$  decadencia y extinción. El análisis cualitativo de este modelo ha sido estudiado en [Aracil *et al.* 97].

Este modelo logístico puede ser enriquecido con la inclusión de un retraso en el bucle de realimentación de la variable que crece. La introducción de este retraso incorpora nuevos, interesantes e inesperados fenómenos en el comportamiento del sistema. Estos nuevos comportamientos son similares a los encontrados por Karsky en un modelo ecológico [Karsky92] y que denominó *catástrofe*



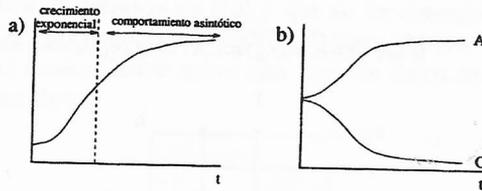


Figura 4: Modelo con crecimiento logístico

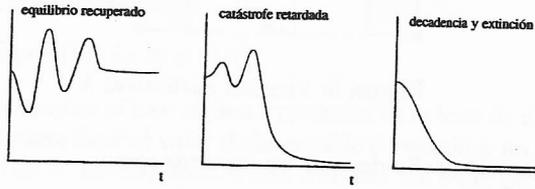


Figura 5: Modelo de crecimiento logístico con retraso

*retardada.*

Este modelo con retraso tiene los dos atractores comentados anteriormente. Karsky estudió el modelo bajo ciertas condiciones iniciales entorno al atractor *A* de crecimiento logístico normal. Perturbando el sistema alrededor de este atractor, el sistema comenzaba a oscilar para posteriormente recuperar el equilibrio. A este comportamiento le denominó *equilibrio recuperado* (figura 5.a).

Por otro lado, cuando esta perturbación de los parámetros era grande, el sistema directamente iba al comportamiento que se denominó *decadencia y extinción*, ya que directamente el sistema iba al declive, es decir, al otro atractor del comportamiento bimodal (figura 5.c).

Más aún, el observó un fenómeno curioso, de manera que para variaciones intermedias en los parámetros, el sistema inicialmente oscilaba, pero finalmente caía en el declive. A este comportamiento lo denominó *catástrofe retardada* (figura 5.b). La catástrofe ocurre aunque inicialmente parece que no iba a ocurrir.

Sea *S* el modelo semicualitativo logístico con restricciones cuyas ecuaciones diferenciales se describen a continuación donde se ha añadido un retraso en la realimentación, quedando:

$$\Phi \equiv \begin{cases} \dot{x} = x(nr - m), & y = \text{retraso}_\tau(x), \\ x > 0, & r = h(y), \\ h \equiv \{(-\infty, -\infty), +, (d_0, 0), +, (0, 1), +, (d_1, e_0), -, (1, 0), -(+\infty, -\infty)\} \end{cases}$$

En este modelo, *n* es el factor de crecimiento, *m* es el factor de decrecimiento, *h* una función cualitativa con un máximo en  $(x_1, y_0)$  cuya representación gráfica se recoge en la (figura 6).

Este modelo responde a la definición de modelo semicualitativo con restricciones. Por un lado tiene conocimiento cualitativo que viene determinado por la función *h*, existe una variable intermedia *r* y además tiene la restricción  $x > 0$ .

Las condiciones iniciales del sistema se describen mediante una serie de operadores unarios cuali-

$$h \equiv \{(-\omega, -\omega), +(d_0, 0), +(0, 1), +(d_1, e_0), -(+\omega, -\omega)\}$$

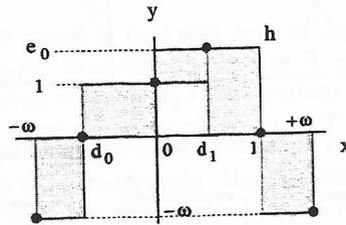


Figura 6: Función cualitativa  $h$

tativos, cuyos valores han sido definidos por los expertos como:

$$\Phi_0 \equiv \begin{cases} X_0 \in PM_x, \\ PP_x(m), PP_x(n), \\ \tau \in PM_\tau, PG_\tau \end{cases}$$

siendo  $PP, PM, PG$  operadores unarios cualitativos para las variables  $x, \tau$ , definidos como:

$$\langle x, 1, 3, 50 \rangle, \quad \langle \tau, 1, 2, 10 \rangle$$

donde los parámetros  $n, m$  toman valores pertenecientes a los operadores unarios definidos para la variable  $x$ . Según los expertos, los valores de infinito para estas magnitudes son  $\infty_x = 100$  y  $\infty_\tau = 10$ .

Sobre este sistema interesa conocer lo siguiente:

1. Si siempre se alcanza un equilibrio estable.
2. Si en estos equilibrios, hay alguno cuyo valor es distinto de cero, esto es se alcanza un equilibrio distinto al de la decadencia.
3. Si todas las trayectorias que alcanzan el equilibrio de decadencia, lo hacen sin oscilar.
4. Clasificar la base de datos de acuerdo con los patrones de comportamientos del sistema.

Mediante la aplicación de la metodología se responderá a las consultas y se clasificará la base de datos de acuerdo con los comportamientos que se encuentren.

En primer término, se aplica a  $S$  las reglas de transformación de manera que se transforme  $S$  en  $F$ .

$$\begin{cases} \dot{x} = x(n\tau - m), & y = \text{retraso}_\tau(x), \\ x > 0, & \tau = \Psi_H(y), \\ x_0 \in [1, 3], & m, n \in [0, 1], \quad \tau \in [1, 10] \end{cases} \quad (10)$$

Los intervalos de cada una de las variables de estado y parámetros se definen de acuerdo con el conocimiento de los expertos, aplicando las técnicas de transformación sobre los operadores unarios suministrado en las condiciones iniciales del sistema. Por otro lado, la función cualitativa  $h$  se transforma en el conjunto de restricciones  $\Psi_H$  como se describe en [Ortega00].

Sobre la familia de modelos  $F$  definida en (10) y que se ha obtenido tras aplicar las técnicas de transformación se aplica el algoritmo **GenerarBD( $F$ )** para obtener la base de datos de trayectorias. Las consultas que se desean hacer sobre esta base de datos se formulan con el lenguaje de consulta/clasificación como sigue:

$$1. \forall r \in T \bullet EQ$$

Con esta instrucción estamos indicando si todas las trayectorias almacenadas en la base de datos acaban en equilibrio estable. La respuesta a esta consulta fue *true*, por lo que no hay ciclos límites en el comportamiento del sistema.

$$2. \exists r \in T \bullet (EQ \wedge \text{siempre } (t \simeq t_f \Rightarrow x \neq 0))$$

Esta consulta permite descubrir si hay alguna trayectoria en la base de datos que acabe en equilibrio estable y además en la parte final el valor de la variable de estado  $x$  no esté próximo a cero, siendo  $\approx$  un operador binario de  $x$ . La respuesta a esta consulta fue *true*, por lo tanto se ha descubierto un posible comportamiento del sistema que corresponde con el que en [Karsky92] se denomina *equilibrio recuperado*.

$$3. \forall r \in T \bullet (EQ \wedge \text{siempre } (t \simeq t_f \Rightarrow x \approx 0) \wedge \text{longitud}([\dot{x} = 0], \{x\}) = 0)$$

En esta tercera consulta interesa conocer si todas las trayectorias que acaban en un equilibrio estable y el valor de la variable de estado al final de la simulación es cercano a 0 ( $x \approx 0$ ) lo hacen sin oscilar. En una trayectoria hay oscilaciones si en la misma no hay ni máximos ni mínimos. La lista de puntos donde  $\dot{x} = 0$  es la lista con los máximos y mínimos relativos de una trayectoria. Cuando la longitud de esta lista es 0 indica que no hay oscilaciones en la misma al carecer de máximos y por lo tanto de mínimos.

La respuesta a esta consulta fue *false*, es decir, hay trayectorias que alcanzan el comportamiento cercano a cero (catástrofe) oscilando y sin oscilar. Es decir, se han descubierto los otros dos comportamientos que Karsky indicaba en su trabajo.

4. Por lo tanto la base de datos de trayectorias se va a clasificar con las etiquetas que representan a los tres comportamientos descubiertos: *equilibrio recuperado*, *catástrofe retardada* y *decadencia y extinción*. Esta clasificación utilizando el lenguaje se propone como:

$$[r, EQ \wedge \text{siempre } t \simeq t_f \Rightarrow x \neq 0] \Rightarrow \text{equilibrio recuperado}$$

$$[r, EQ \wedge \text{siempre } (t \simeq t_f \Rightarrow x \approx 0) \wedge \text{longitud}([\dot{x} = 0], \{x\}) > 0] \Rightarrow \text{catástrofe retardada}$$

$$[r, EQ \wedge \text{siempre } (t \simeq t_f \Rightarrow x \approx 0) \wedge \text{longitud}([\dot{x} = 0], \{x\}) = 0] \Rightarrow \text{decadencia y extinción}$$

Al aplicar sobre la base de datos  $T$  las anteriores reglas de clasificación, provocaron que todas las trayectorias de  $T$  se clasificaran con una y solo una de las etiquetas anteriores.

De forma análoga, al aplicar sobre la base de datos original  $T$  un algoritmo de *clustering*, se encontraron los tres comportamientos del sistema tomando el grado de afinidad como 0.05<sup>1</sup>. Además las trayectorias que forman cada uno de los *clusters* se agrupaban de manera análoga a como lo hace cuando se aplican las anteriores reglas de clasificación. Por lo tanto, cada grupo de *clusters* encontrado corresponde con uno de los posibles comportamientos del sistema.

<sup>1</sup>Un cinco por ciento respecto a la distancia media entre trayectorias.

Tomando como entrada los parámetros del sistema  $m, n, \tau$  y el valor la variable  $x$  en el instante inicial de la simulación  $x_0$  se aplicaron los algoritmos que obtienen los sistemas de reglas.

Con el C4.5 [Quinlan93] se obtuvo un árbol de decisión cuyo sistema de reglas lo forman 17 reglas y un error del 5%. La herramienta C4.5 también proporciona un árbol podado a partir del anterior. El que se obtuvo para este caso generaba un sistema de 15 reglas pero el error era del 6.5%. Centrando el estudio solamente sobre los parámetros del sistema  $m, n, \tau$  el sistema resultante también tenía 18 reglas y un error del 6% para la muestra, y su árbol podado con 16 reglas y un error del 5.2%

Al aplicar la herramienta COGITO [Aguilar97] a la base de datos se obtuvieron varios sistemas de reglas, unos con mayor número de reglas y menor error, y otros al contrario. Algunos de estos sistemas obtuvieron para el caso de la clasificación según  $m, n, \tau, x_0$  un sistema con 12 reglas y un error del 7.8%, 8 reglas y un error del 5.8%, o un sistema con 7 reglas y un error del 5.2%.

En el apéndice A de este artículo se recogen los sistemas de reglas que se obtuvieron al aplicar las herramientas C4.5 y COGITO.

Los resultados obtenidos con la metodología propuesta en esta tesis están en concordancia con los obtenidos por Karsky en su trabajo y que fueron formalmente demostrados mediante un razonamiento matemático por Aracil en su trabajo [Aracil *et al.* 97], donde además se hace un estudio bastante pormenorizado de este sistema.

## 7 Conclusiones y trabajo futuro

En este artículo se ha presentado de manera resumida una nueva metodología para automatizar el análisis de sistemas dinámicos con conocimiento cuantitativo y cualitativo. Esta metodología se basa en un proceso de la transformación, la aplicación de técnicas estocásticas, simulación cuantitativa, generación de una base de datos de trayectorias y definición de un lenguaje de consulta/clasificación. La simulación se realiza por medio de las técnicas estocásticas y los resultados se guardan en una base de datos cuantitativa, que se puede clasificar con el lenguaje propuesto. Una vez la base de datos se clasifica, se pueden aplicar algoritmos genéticos para obtener conclusiones sobre el sistema dinámico.

En el futuro, nosotros vamos a enriquecer el lenguaje propuesto con: operadores para comparar trayectorias entre sí, lógica temporal entre tiempos distintos de una misma trayectoria, más tipo de ecuaciones,...

Una Compañía metalúrgica está interesada en modificar su sistema de producción de acero aplicando la metodología propuesta. En futuros trabajos, describiremos este sistema en detalle y las conclusiones obtenidas con la aplicación de la metodología propuesta.

### Apéndice A

La aplicación de C4.5 sobre la base de datos formada por 200 trayectorias, obtuvo el siguiente árbol de decisión que depende de los parámetros del sistema  $m, n, \tau$  y del valor la variable  $x$  en el instante inicial de la simulación:

```

M <= 0.170079 :
| N > 0.169506 : DecaeRetardada (17.0/2.0)
| N <= 0.169506 :
| | N <= 0.0845729 : DecliveExtinción (6.0/1.0)
| | N > 0.0845729 : EquilibrioRecuperado (2.0)
M > 0.170079 :
| M <= 0.511779 :
| | N <= 0.405235 :
| | | N <= 0.287256 : DecliveExtinción (20.0)
| | | N > 0.287256 :
| | | | X0 <= 1.87752 : EquilibrioRecuperado (2.0)
| | | | X0 > 1.87752 : DecliveExtinción (5.0)
| | | N > 0.405235 :
| | | | X0 <= 1.64469 :
| | | | | M > 0.270087 : DecaeRetardada (5.0)
| | | | | M <= 0.270087 :
| | | | | | M <= 0.231029 : DecaeRetardada (3.0)
| | | | | | M > 0.231029 : EquilibrioRecuperado (3.0)
| | | | | X0 > 1.64469 :
| | | | | | N > 0.810159 : DecliveExtinción (2.0/1.0)
| | | | | | N <= 0.810159 :
| | | | | | | N <= 0.42631 : DecliveExtinción (2.0)
| | | | | | | N > 0.42631 :
| | | | | | | | TAU <= 9.52249 : EquilibrioRecuperado (9.0/1.0)
| | | | | | | | TAU > 9.52249 : DecliveExtinción (2.0)
| M > 0.511779 :
| | X0 <= 1.22575 : DecliveExtinción (3.0/1.0)
| | X0 > 1.22575 :
| | | N <= 0.792701 : DecliveExtinción (55.0/1.0)
| | | N > 0.792701 :
| | | | TAU <= 2.56321 : EquilibrioRecuperado (3.0)
| | | | TAU > 2.56321 : DecliveExtinción (14.0/1.0)

```

Este árbol genera un sistema con 17 reglas y un error del 5%. Esto quiere decir que de cada cien trayectorias de la base de datos al aplicar la anterior clasificación noventa y cinco se clasificarían con su etiqueta y cinco con otra que no es la que el algoritmo de *clustering* le había asignado.

Las reglas que se obtienen son:

```

si M <= 0.170079 y N > 0.169506
  entonces DecaeRetardada
si M <= 0.170079 y N <= 0.0845729
  entonces DecliveExtinción
si M <= 0.170079 y 0.0845729 < N <= 0.169506
  entonces EquilibrioRecuperado
si 0.170079 < M <= 0.511779 y N <= 0.287256
  entonces DecliveExtinción
si 0.170079 < M <= 0.511779 y 0.287256 < N <= 0.405235 y X0 <= 1.87752

```

entonces EquilibrioRecuperado  
 si  $0.170079 < M \leq 0.511779$  y  $0.287256 < N \leq 0.405235$  y  $X_0 > 1.87752$   
 entonces DecliveExtinción  
 si  $0.270087 < M \leq 0.511779$  y  $0.405235 < N$  y  $X_0 \leq 1.64469$   
 entonces DecaeRetardada  
 si  $0.170079 < M \leq 0.231029$  y  $0.405235 < N$  y  $X_0 \leq 1.64469$   
 entonces DecaeRetardada  
 si  $0.231029 < M \leq 0.270087$  y  $0.405235 < N$  y  $X_0 \leq 1.64469$   
 entonces EquilibrioRecuperado  
 si  $0.170079 < M \leq 0.511779$  y  $0.810159 < N$  y  $X_0 > 1.64469$   
 entonces DecliveExtinción  
 si  $0.231029 < M \leq 0.270087$  y  $0.405235 < N \leq 0.42631$  y  $X_0 > 1.64469$   
 entonces DecliveExtinción  
 si  $0.231029 < M \leq 0.270087$  y  $0.42631 < N \leq 0.810159$  y  $X_0 > 1.64469$  y  $TAU \leq 9.52249$   
 entonces EquilibrioRecuperado  
 si  $0.231029 < M \leq 0.270087$  y  $0.42631 < N \leq 0.810159$  y  $X_0 > 1.64469$  y  $TAU > 9.52249$   
 entonces DecliveExtinción  
 si  $M > 0.511779$  y  $X_0 \leq 1.22575$   
 entonces DecliveExtinción  
 si  $M > 0.511779$  y  $X_0 > 1.22575$  y  $N \leq 0.792701$   
 entonces DecliveExtinción  
 si  $M > 0.511779$  y  $X_0 > 1.22575$  y  $N > 0.792701$  y  $TAU \leq 2.56321$   
 entonces EquilibrioRecuperado  
 si  $M > 0.511779$  y  $X_0 > 1.22575$  y  $N > 0.792701$  y  $TAU > 2.56321$   
 entonces DecliveExtinción

La herramienta *C4.5* también proporciona un árbol podado a partir del anterior. El que se obtuvo para este caso generaba un sistema de 15 reglas pero el error era del 6.5%.

Centrando el estudio solamente sobre los parámetros del sistema  $m, n, \tau$  el sistema resultante también tenía 18 reglas y un error del 6% para la muestra, y su árbol podado con 16 reglas y un error del 5.2%.

Al aplicar la herramienta *COGITO* a la base de datos se obtuvieron varios sistemas de reglas. Algunos de estos sistemas obtuvieron para el caso de la clasificación según  $m, n, \tau, x_0$  un sistema con: 12 reglas y un error del 7.8%, 8 reglas y un error del 5.8%, o un sistema con 7 reglas y un error del 5.2%, siendo este último el que se muestra a continuación:

si  $0.304787 < M$  y  $N < 0.692619$  y  $1.012669 < X_0$   
 entonces DecliveExtinción  
 sino si  $0.512724 < M$  y  $1.311880 < TAU$   
 entonces DecliveExtinción  
 sino si  $0.080025 < N$  y  $4.162784 < TAU$  y  $X_0 < 1.705117$   
 entonces DecaeRetardada  
 sino si  $N < 0.313170$  y  $1.336671 < TAU$  y  $1.059963 < X_0$   
 entonces DecliveExtinción  
 sino si  $0.070686 < M$  y  $N < 0.835682$  y  $TAU < 7.466669$   
 entonces EquilibrioRecuperado

sino si  $0.192920 < M$  y  $N < 0.991590$  y  $1.492503 < \text{TAU}$   
 entonces DecliveExtinción  
 sino si  $\text{TAU} < 9.038445$  y  $1.057524 < X0$   
 entonces DecaeRetardada

Este sistema podría equipararse al obtenido con la herramienta C4.5 aunque el número de reglas que se obtiene con *COGITO* es menor, si bien las reglas de esta última herramienta son jerárquicas.

Cuando solamente se tienen en cuenta los parámetros del sistema  $m, n, \tau$  se presentan de manera resumida algunos de los resultados que se obtuvieron: 13 reglas y error del 2.6%, 9 reglas y error del 5.8%, 6 reglas y error del 8.1%, este sistema es el que se muestra a continuación:

si  $0.437458 < M$  y  $N < 0.793310$   
 entonces DecliveExtinción  
 sino si  $0.080413 < M$  y  $N < 0.383127$   
 entonces DecliveExtinción  
 sino si  $M < 0.526804$  y  $0.475383 < N$  y  $2.170260 < \text{TAU} < 9.517801$   
 entonces DecaeRetardada  
 sino si  $0.045368 < M < 0.831606$  y  $N < 0.869349$  y  $\text{TAU} < 6.125667$   
 entonces EquilibrioRecuperado  
 sino si  $0.193584 < M$  y  $2.446866 < \text{TAU}$   
 entonces DecliveExtinción  
 sino si  $M < 0.681996$  y  $0.058356 < N$   
 entonces DecaeRetardada

## Referencias

- [Aguilar97] Aguilar J., Riquelme J. y Toro M. "COGITO: Un sistema de autoaprendizaje basado en algoritmos genéticos" III Jornadas de Informática, El Puerto de Santa María (Cádiz), pág 79-88 (1997).
- [Aracil et al. 97] J. Aracil, E. Ponce, L. Pizarro. Behavior patterns of logistic models with a delay *Mathematics and computer in simulation* 44: 123-141, 1997.
- [Gasca98] R.M. Gasca. Razonamiento y Simulación en Sistemas que integran conocimiento cualitativo y cuantitativo. *Ph.D. Thesis*, University of Sevilla (Spain), 1998.
- [Karsky92] karsky M., Dore J.C. y Gueneau P. "Da la possibilité d'apparition de catastrophes différés" *Ecodecision* n° 6, Montreal (Canadá), 1992.
- [Kay96] H. Kay. Refining imprecise models and their behaviors. Ph.D. Thesis University of Texas (USA), 1996.
- [Kuipers94] B.J. Kuipers. *Qualitative reasoning. Modeling and simulation with incomplete knowledge*, The MIT Press, Cambridge, Massachusetts 1994.
- [Ortega et al. 97] J.A. Ortega, R.M. Gasca, M. Toro. Automatización del análisis y simulación de modelos cualitativos mediante restricciones. En *Proc VII Conferencia Asociación Española para Inteligencia Artificial*, pages 85-94, Málaga (Spain), November 1997.



- [Ortega *et al.* 98] J.A. Ortega, R.M. Gasca, M. Toro. Including qualitative knowledge in semiquantitative dynamical systems. *Lecture Notes in Artificial Intelligence* 1415: 329–335, 1998.
- [Ortega *et al.* 99] J.A. Ortega, R.M. Gasca, M. Toro. Behavior patterns of semiquantitative dynamic systems by means of quantitative simulations *Proceedings International Joint Conference on Artificial Intelligence 1999* Estocolmo (Suecia) Agosto-1999. (por aparecer)
- [Ortega00] J.A. Ortega. Patrones de comportamiento temporal en modelos semicualitativos con restricciones. *Ph.D. Thesis*, University of Sevilla (Spain), 2000.
- [Quinlan93] Quinlan J. “*C4.5: programs for machine learning*” Morgan Kauffmann Publisher, Inc., 1993.
- [Richardson95] Richardson G. “*Loop polarity, loop dominance, and the concept of dominant polarity*” *Systems dynamics Review* n° 11 (1), pág. 67-88 (1995).
- [Travé-Massuyès *et al.* 97] L. Travé-Massuyès, Ph. Dague, F. Guerrin. *Le raisonnement qualitatif pour les sciences de l'ingénieur*. Hermes Ed., Paris 1997.
- [Vescovi *et al.* 95] M. Vescovi, A. Farquhar, and Y. Iwasaki. Numerical interval simulation: combined qualitative and quantitative simulation to bound behaviors of non-monotonic systems. *Proc. 14th Int. Joint Conference on Artificial Intelligence* pages 1806–1812, Canada, 1995.

