

Similaridad para Sistemas Dinámicos con Base en Criterios Característicos de sus Trayectorias

Hernando Castañeda Marín

UNIVERSIDAD DE PAMPLONA.

UNIVERSIDAD DE LOS ANDES, Doctorado en Ciencias Aplicadas, Mérida, Venezuela
hcastaneda@unipamplona.edu.co

Eliézer Colina Morales

UNIVERSIDAD DE LOS ANDES, Departamento de Sistemas de Control, Mérida, Venezuela
ecolina@ula.ve

Wladimir Rodríguez Graterol

UNIVERSIDAD DE LOS ANDES, Departamento de Computación, Mérida, Venezuela
wladimir@ula.ve

Recibido para revisión May-2006, aceptado Jun-2006, versión final recibida Jun-2006

Resumen: Este trabajo se define y se propone un procedimiento para la determinación de la similaridad entre dos trayectorias que representan los comportamientos temporales de sus respectivos sistemas dinámicos complejos. Este tipo de aplicaciones es una necesidad frecuente en el área de minería de datos, donde es frecuente encontrar problemas para determinar la similaridad entre dos series de tiempo. El procedimiento propuesto está basado en el uso de diferentes funciones de escala y traslación y se hace uso de la teoría de conjuntos difusos a objeto de modelar una representación gradual de la similaridad, incorporando elementos análogos al juicio humano y métodos de la geometría computacional. El trabajo contempla algoritmos para computar la noción considerada de similaridad para casos de trayectorias unidimensionales y multidimensionales. Igualmente se determina la complejidad temporal de dichos algoritmos a partir de aplicaciones simuladas de sistemas dinámicos.

Palabras Clave: Modelamiento e identificación de sistemas, sistemas dinámicos, Inteligencia artificial.

Abstract: This work proposes two fuzzy logic based procedures for determining similarities between complex temporal dynamic system trajectories. These types of problems may be often found in the data mining area, where it is important to establish similarities between time series. The proposed procedure is based upon using different scale and translation functions and uses the theory of fuzzy sets in order to model a gradual representation of similarities, incorporating analog elements to human judgment and geometrical computational methods. The work contemplates algorithms for computing a similarity notion for both one-dimensional and multidimensional trajectories. Finally, the work includes some computer simulation examples in order to determine the temporal complexity of the proposed algorithms.

Keywords: Modeling and Identification Systems, Dynamic System, Artificial intelligent.

1 INTRODUCCIÓN

Las trayectorias asociadas a los comportamientos de sistemas dinámicos, particularmente las correspondientes a series de tiempo, constituyen una importante clase compleja de datos que aparecen en muchas aplicaciones financieras, médicas y de ingeniería de procesos. Aun cuando existe una amplia literatura técnica dedicada al análisis de series de tiempo, que da respaldo conceptual

a la resolución de problemas, como los de predicción, la noción de similaridad para series de tiempo, en aplicaciones de análisis de datos, no ha sido suficientemente estudiada.

La similaridad entre trayectorias puede ser interpretada en forma diferente dependiendo del contexto. En el lenguaje natural, la interpretación de similaridad está asociada con "tener características comunes" o "no tener diferencia en la forma pero sí en tamaño o posición"

[Setnes y Kaymak (1998), p. 378].

De allí que, intuitivamente dos trayectorias pueden ser consideradas similares si exhiben comportamiento similar, a lo largo de una parte grande de su longitud. Así, para definir matemáticamente similaridad resulta esencial determinar cuales propiedades de las trayectorias pueden ser utilizadas, de modo de describir su comportamiento y en que forma la comparación (matching) de las trayectorias puede ser llevada a cabo.

En ciertas aplicaciones, el propósito de la clasificación de las series de tiempo es la partición de las mismas en grupos o series con dinámica similar. En estos casos la noción de similaridad es utilizada para cuantificar la aproximación entre sistemas dinámicos y sus atractores, más que como series de tiempo individuales. Para sistemas dinámicos con grados de libertad los atractores son definidos como un subconjunto M -dimensional en el espacio de fase hacia los cuales algunas de las trayectorias se juntan como "atraídas" asintóticamente.

2 NOCIONES DE SIMILARIDAD ESTRUCTURAL PARA TRAYECTORIAS DE SISTEMAS DINÁMICOS

Para determinar la similaridad estructural son analizados aspectos relevantes del comportamiento de las trayectorias, que definen propiedades matemáticas de las mismas como pendiente, curvatura, posición, valores de puntos de inflexión y suavidad, los cuales son utilizados como criterios de comparación. De esta forma, la similaridad estructural es ajustada a situaciones en la que se determina patrones particulares en las trayectorias que pueden emparejarse. De esta manera, la tarea es definir una medida de similaridad para trayectorias, que expresen un grado de emparejamiento de acuerdo a algún criterio predefinido e invariante, como Cambios de escala, traslación, observaciones lejanas o apariencia de valores incorrectos.

2.1 Medida de Similaridad Estructural

Establecer similaridad estructural entre trayectorias, significa hacer una búsqueda para medir el comportamiento temporal similar entre las mismas. El punto crucial es sin embargo, ensamblar un concepto de similaridad que sea de fácil manejo computacional. Algunas veces el comportamiento estadístico global es importante, mientras que en otros casos, los cambios locales en el comportamiento son más esenciales. Así, es difícil definir una medida de similaridad específica que pueda ser útil en todas las aplicaciones. La naturaleza diferente de las trayectorias requiere de la selección específica de una medida adecuada de similaridad. En consecuencia, parece sensato intentar definir un esquema general para la medida de similaridad; para luego, dependiendo de cada

caso específico, determinar cuales aspectos son más importantes para el análisis de trayectorias.

Así, la similaridad estructural puede ser vista como un grado de aproximación de las trayectorias con respecto a aspectos como forma, evolución, tendencia, tamaño u orientación en, que permiten representar su comportamiento. Dependiendo del aspecto seleccionado, diferentes propiedades matemáticas y características de las trayectorias pueden ser relevantes para describir la similaridad; por ejemplo, la pendiente, curvatura, posiciones, valores de puntos extremos u otras informaciones como suavidad y monotonía. Un conjunto de características relevantes tienen que ser escogidas de tal forma que las trayectorias se aproximen una con respecto a la otra, denotando características que presentan comportamientos similares.

2.2 Definición de Similaridad Estructural Basada en la Tendencia de las Trayectorias

La tendencia se define como un cambio lento pero consistente de una trayectoria en forma unidireccional que describe su comportamiento general. La forma de tendencia puede ser por ejemplo la pendiente, un paso o un paso amortiguado. La tendencia parece similar a una pendiente que puede ser estimada mediante el ajuste de una línea recta para un conjunto de mediciones dadas. (constituye una trayectoria), con la aproximación de los mínimos cuadrados.

Dado un conjunto de mediciones donde $(x_1, t_1), (x_2, t_2), \dots, (x_p, t_p)$ son conocidos y están sujetas a errores de mediciones. La tarea de aproximación consiste en estimar mediante una línea de regresión $f(t) = a_1 t + a_2$. Los coeficientes de regresión a_1 y a_2 son determinados de modo que el valor $\hat{x}_k = a_1 t_k + a_2, k = 1, \dots, p$ sean los próximos más posibles a los valores en el sentido de mínimo cuadrado.

Esto requiere la minimización de los errores de la suma de los cuadrados:

$$e^2 = \sum_{k=1}^p (x_k - \hat{x}_k)^2 \quad (1)$$

Los coeficientes de regresión pueden también ser estimados, utilizando la covariancia de y [Ruger (1989)], así:

$$a_1 = \frac{\text{cov}(x, t)}{\text{var}(t)} \quad a_2 = \bar{x} - a_1 \bar{t} \quad (2)$$

Donde:

$$\text{cov}(x, t) = \frac{1}{p} \sum_{k=1}^p (x_k - \bar{x})(t_k - \bar{t}) \quad (3)$$

$$\text{var}(t) = \sum_{k=1}^p (t_k - \bar{t})^2 = \sum_{k=1}^p (t_k^2 - \bar{t}^2) \quad (4)$$

\bar{t} y \bar{x} son los valores medios de t y $x(t)$ sobre el intervalo de tiempo $[t_1, t_p]$.

Definición 1. Dos trayectorias son consideradas similares con respecto a su tendencia temporal si están caracterizadas mediante valores similares del parámetro.

Definición 2. Dos trayectorias son consideradas similares, con respecto a su tendencia temporal y localización si están caracterizadas por valores similares para los parámetros a_1 y a_2 .

2.3 Definición de la Similaridad en base a la Curvatura de las Trayectorias

La curvatura de las trayectorias de cada punto describe el grado con la cual una trayectoria está doblada en este punto. Esta es evaluada mediante el coeficiente de la segunda derivada de una trayectoria en cada punto, que puede ser definida mediante la siguiente ecuación (para una trayectoria unidimensional)

$$cv_k = x_k'' = \frac{x_k' - x_{k-1}'}{t_k - t_{k-1}} \quad k = 3, \dots, p \quad (5)$$

Donde x_k' denota el coeficiente de la primera derivada en el punto x_k dado por:

$$x_k' = \frac{x_k - x_{k-1}}{t_k - t_{k-1}} \quad k = 2, \dots, p \quad (6)$$

Sustituyendo la anterior ecuación en la ecuación de la curvatura, se llega a la siguiente ecuación basada en los valores de las trayectorias originales:

$$cv_k = x_k'' = \frac{x_k - 2x_{k-1} + x_{k-2}}{(t_k - t_{k-1})^2} \quad k = 3, \dots, p \quad (7)$$

Si las trayectorias poseen mínimos y máximos locales, éstos pueden ser detectados mediante la espera de un cambio de signo en los valores de la primera derivada. Así, es suficiente considerar los coeficientes de la segunda derivada ya que en estos puntos específicos es donde la curvatura es máxima. La característica distintiva cuando se considera la curvatura, es el signo del coeficiente de la segunda derivada. Si el coeficiente es positivo en cierto periodo de tiempo, entonces la trayectoria es convexa en el intervalo (cerca al tope). Si el coeficiente es negativo en determinado periodo de tiempo, la trayectoria es cóncava (cerca al punto bajo). Si el coeficiente es igual a cero en algún punto, que es llamado punto de inflexión, no se presenta curvatura en este punto. Los puntos de inflexión aparecen en trayectorias oscilatorias e indica el cambio de curvatura de convexo a cóncavo y viceversa. Todas las funciones lineales son caracterizadas mediante curvatura cero en todos los puntos.

Definición 3. Dos trayectorias son consideradas similares con respecto a su curvatura si están caracterizadas por coeficientes similares en su segunda derivada.

2.4 Definición de Similaridad basada en la Suavidad de las Trayectorias

La suavidad de una trayectoria describe el grado de oscilación en el comportamiento y puede ser caracterizado por el número de cambios de signo en su segunda derivada. Para estimar la suavidad de una trayectoria, se considera un vector de coeficientes de la segunda derivada de la trayectoria $x(t) = [x_1, \dots, x_p]$ y se transforma en un vector binario, utilizando las siguientes reglas: valores positivos son sustituidos por +1, valores negativos son sustituidos por -1 y valores cero son no considerados. El resultado es el vector $h = [h_1, \dots, h_{(p-z-2)}]$, donde z es el número de valores de cero en las segundas derivadas. Se determina un segundo vector w cuyos elementos w_k , $k = 1, \dots, p-z-3$ corresponden a sumas de cada par de elementos sucesivos h_k y h_{k+1} . Los elementos w_k toman sus valores del conjunto $\{-2, 0, 2\}$, donde el valor de cero corresponde al cambio de signo en el vector de segundas derivadas. De esta manera, determinando el número de valores de cero en el vector w , el número de cambios de signo de las segundas derivadas sm se calcula así:

$$sm = \sum_{k=1}^{p-z-3} w_k = \sum_{k=1}^{p-z-3} (h_k + h_{k+1}) \quad \forall w_k = 0 \quad (8)$$

Definición 4. Dos trayectorias $x(t)$ e $y(t)$ son consideradas similares con respecto a su suavidad si existen características de valores similares del parámetro sm .

2.5 Definición de Similaridad Estructural basada en Parámetros Temporales Específicos de las Trayectorias.

Para algunos problemas, además de las medidas de similaridad basadas en las características de la curvatura y suavidad en el comportamiento general de las trayectorias con respecto a su forma y característica oscilatoria, puede ser importante considerar parámetros concretos de simples ondulaciones que aparecen en las trayectorias para identificar patrones temporales similares en las mismas. Los patrones temporales en la trayectoria se pueden descomponer en segmentos indicando las tendencias locales; de tal forma que cada segmento está limitado mediante puntos de inflexión o por un punto de inflexión y un punto extremo t , [Bakshi, Locher y Stephanopoulos (1994)].

Este conjunto de características proporciona la información necesaria acerca del desarrollo temporal de una trayectoria a manera de pedazos, pero se puede ampliar con muchos más parámetros para una más completa

descripción de las trayectorias. Las siguientes características temporales pueden ser consideradas adicionalmente:

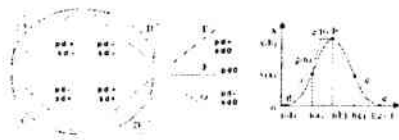


Figura 1: Características temporales cualitativas y cuantitativas obtenidas por segmentación

- $gg(b)$ es el valor de la curvatura en el extremo, wd es la duración de un patrón, como una colina, que es definida con respecto al punto de inflexión (duración de segmentos elementales) o con respecto a la línea de principio del patrón (duración de cuatro segmentos).
- tn_1 es el intervalo de tiempo hasta el primer valor cero de la trayectoria, tn_2 es el intervalo de tiempo hasta el segundo valor cero de la trayectoria.
- I es una integral de la parte de la trayectoria hasta el primer valor de cero. CG es el centro de gravedad de la parte de la trayectoria hasta el primer valor de cero.
- Ald es el valor de la mediana de los valores de las trayectorias. RV es el rango de los valores de las trayectorias.
- IV es la platea o el valor límite de la trayectoria. Medidas estadísticas (ej. Valores de media, desviación estándar, coeficiente de correlación entre segmentos de trayectorias).

Definición 5. Dado un conjunto de parámetros temporales pertinentes, las trayectorias $x(t)$ e $y(t)$ pueden ser consideradas similares si los valores de esos parámetros describen patrones elementales similares en las mismas.

3 RESULTADOS EXPERIMENTALES

A continuación se ilustra el comportamiento de un sistema dinámico, específicamente el movimiento de un péndulo simple, donde se observa el comportamiento de las variables de estado, velocidad angular y posición angular a través de intervalos de tiempo muy pequeños. Se quiere analizar los rasgos del comportamiento de las variables de estado en términos de similitud estructural entre dos trayectorias.

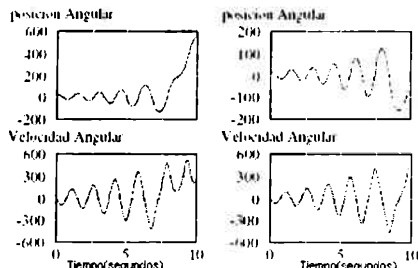






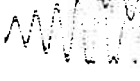
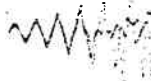






Figura 2: Comportamiento de las variables de estado de un péndulo simple para dos diferentes posiciones angulares

La definición de similitud estructural se ilustra en la Tabla 1 donde las trayectorias presentan comportamientos decremental en la posición angular e incremental en la velocidad angular, aunque estas trayectorias son más similares con respecto a su forma, ellas no son similares con respecto a su tendencia temporal. La correspondiente medida de similitud $s(x, y)$ puede ser calculada con el algoritmo 3.1, donde el parámetro a_1 se usa como característica K de la trayectoria y el conjunto difuso A denota "admisible diferencia para la tendencia", esta similitud estructural puede ser aplicada para encontrar clústeres de trayectorias con una tendencia similar, donde la traslación de las trayectorias en el espacio característico a lo largo de las M dimensiones o traslación en el tiempo y grado de sus fluctuaciones son irrelevantes.

Si la tendencia de las trayectorias es un criterio irrelevante por análisis, por ejemplo la localización de las trayectorias en el espacio característico tiene que ser considerada junto con la tendencia específicamente, y serán similares los valores de los parámetros a_1 y a_2 .

La medida de similitud $S(x, y)$ puede ser calculada con el algoritmo 3.1 o una de sus extensiones 3.2, donde la característica K es un vector de coeficientes de la segunda derivada $K = [cc_3, cc_4, \dots, cc_p]$ el conjunto difuso A denota "diferencia admisible para la curvatura" si la traslación temporal es irrelevante para el proceso de reconocimiento de patrones similares en las trayectorias, los vectores de características obtenidas para ambas trayectorias pueden cambiarse cíclicamente respecto de una con la otra y la medida de similitud se define para cada combinación; en esta forma la máxima similitud corresponde al mejor emparejamiento de las trayectorias con respecto a la curvatura encontrada. La medida de similitud basada en la curvatura esta particularmente disponible para trayectorias con bajo número de fluctuaciones y en forma ondulada; esta medida, sin embargo, es sensitiva al cambio de escala, por ejemplo las trayectorias transformadas mediante un cambio en el factor de escala, tienen diferente curvatura.

Tabla 1: Rasgos y grados de membresía

Rasgo de la trayectoria	Parámetros en la variación de estado posición angular	Parámetros en la variación de estado velocidad angular	Medida de similitud en posición angular	Medida de similitud en velocidad angular
Rasgo Tendencia de la trayectoria $O(n)$	 $a = 0.0878$ $a = -12.2035$	 $a = 0.0127$ $a = 1.3690$ $a = 0.0930$ $a = 0.9687$		
Curvatura de las trayectorias $O(n)$				
Suavidad de las trayectorias $O(n)$			$SM = 6$ $SM = 7$ $SM = -1$ Grado de membresía = 0.9974	$SM = 8$ $SM = 8$ $SM = 0$ Grado de membresía = 1
Períodos temporales específicos $O(n)$				

La medida de similitud $s(x, y)$ basada en la suavidad de las trayectorias puede ser obtenida utilizando el algoritmo 3.1 o una de las extensiones 3.2, donde las características K es seleccionado según el parámetro de suavidad sm y el conjunto A difuso denota como "diferencia admisible de suavidad", la medida de similitud puede ser aplicada mediante la comparación de las trayectorias, donde el comportamiento oscilatorio se asume, es más general que la medida de similitud basada en la curvatura puntual de las trayectorias, aunque considerando sólo el valor total de oscilaciones pero no el grado de curvatura de cada onda por separado. Esta medida de similitud está disponible para trayectorias, por las cuales transformaciones de cambios de escala y traslación son irrelevantes; pero es sensitiva a errores de medición. Para evitar la influencia de puntos erróneos en los valores de las similitudes, las trayectorias tienen que estar preprocesadas aplicando técnicas de suavización.

Los parámetros listados en la similitud estructural basada en los parámetros temporales específicos de las trayectorias permiten una descripción de la forma de patrones temporales presentes en las trayectorias: estos tienen en cuenta el número y tamaño de las colinas, su pendiente y curvatura, el instante de su aparición y su duración, donde los factores de cambio de escala y traslación tienen un efecto en los valores de los parámetros, esta medida de similitud está disponible para el reconocimiento y comparación de patrones específicos de trayectorias.

4 CONCLUSIONES

Las anteriores definiciones de medidas de similitud estructural pueden ser utilizadas en diferentes combinaciones para obtener una evaluación más completa de la similitud entre trayectorias. En algunos casos la similitud estructural puede ser reducida a similitud pun-

Tabla 2: Valores de Rasgos en Segmentación

Otros temporales	Trajectory 1 Posición Angular				Trajectory 2 Velocidad Angular				Rasgos Derivados			
	$t(a)$	$x(a)$	$t(b)$	$x(b)$	$t(a)$	$y(a)$	$t(b)$	$y(b)$	Posición angular		velocidad angular	
Primer segmento	$t(a)$	$x(a)$	$t(b)$	$x(b)$	$t(a)$	$y(a)$	$t(b)$	$y(b)$	EE_1	EE_2	EE_3	EE_4
	1.1230	1.1703	14688	22.6828	1.1230	0.9616	14688	30.1014	-380.6	-380.6	-1478	-1478
	2.6230	28446	29688	34.5338	2.6230	0.3280	30000	46.1386	-565.8	-586	-2256	-2256
	4.1230	39633	45000	32.5879	4.1875	3.2387	45025	69.8774	-805.8	-837.3	-3204	-3585
	5.7188	62888	60938	79.5783	5.8436	0.4870	62813	106.3730	-1020	-1021	-4784	-4948
	7.4375	13636	79063	122.044	7.9063	7.4820	85625	178.9490	-931.8	331.3	-144	-2666
Segundo segmento	$t(a)$	$x(a)$	$t(b)$	$x(b)$	$t(a)$	$y(a)$	$t(b)$	$y(b)$	EE_1	EE_2	EE_3	EE_4
	0.0313	15.0000	0.3438	1.8475	0.0313	20.0000	0.3438	2.5908	-109.9	-109.9	-516	-516
	1.5000	22.8662	1.8438	1.2683	1.5000	30.4663	1.8438	3.1972	-140.8	-140.8	-653	-376
	3.0000	34.7554	3.3438	2.1638	3.0813	46.2634	3.3750	3.2782	-215.2	-129.8	-658	-614
	4.5313	52.7599	4.8750	5.2675	4.5938	70.2691	4.9688	5.3384	-349	39.680	-342	-342.3
	6.1250	80.1139	6.5313	8.8229	6.3125	106.9970	6.7500	12.0123	-387.0	821.24	-2379	4286
Tercer segmento	$t(a)$	$x(a)$	$t(b)$	$x(b)$	$t(a)$	$y(a)$	$t(b)$	$y(b)$	EE_1	EE_2	EE_3	EE_4
	0.3750	-0.3472	0.7188	-18.309	0.3750	-0.234	0.7188	24.3618	307.0	306.99	1182	1182
	1.8750	-2.084	2.2188	-28.031	1.8750	-1.215	2.2188	-37.047	467.4	467.35	1835	1835
	3.3750	-2.8751	3.7188	-42.431	3.4063	-3.327	3.7500	-56.124	671.9	696.32	2729	2883
	4.9063	-2.1573	5.2813	-64.652	5.0000	-4.244	5.4063	-86.424	923.4	951.19	3883	4399
	5.2813	-9.9250	6.9088	-98.577	6.7813	-1.030	7.3125	-133.86	1050.0	951.19	573.4	573.4
Cuarto segmento	$t(a)$	$x(a)$	$t(b)$	$x(b)$	$t(a)$	$y(a)$	$t(b)$	$y(b)$	EE_1	EE_2	EE_3	EE_4
	0.7300	-38.534	1.0938	-1.5474	0.7300	-24.6993	1.0938	-26.342	123.21	123.20	471.9	471.9
	2.2300	-28.189	2.3938	-1.7079	2.2500	-37.5436	2.3938	-50.477	167.93	167.92	571.3	571.3
	3.7300	-42.843	4.0938	-5.1353	3.7813	-56.9527	4.1563	-4.7443	301.89	94.363	782.8	478.6
	5.2325	-65.028	5.6875	-7.7172	5.4375	-86.4941	5.8125	-10.7539	360.01	-279.0	602.3	-387.6
	7.0000	-98.808	7.4063	-11.036	7.3438	-133.98	7.8750	-74.106	596.37	-960.5	721.9	-144.5
Quinto segmento	$t(a)$	$x(a)$	$t(b)$	$x(b)$	$t(a)$	$y(a)$	$t(b)$	$y(b)$	EE_1	EE_2	EE_3	EE_4
	9.1875	-160.49	9.8750	-32.889	8.3938	181.7760	9.3125	335.4870				

tual. Por ejemplo, considerando la similitud puntual para la primera derivada de las trayectorias, la comparación de las trayectorias se realiza con respecto a su pendiente puntual; cuando, por el contrario, los parámetros de cambio de escala son ignorados. Si la similitud puntual es determinada mediante la segunda derivada de las trayectorias, entonces la curvatura puntual de las trayectorias es considerada como un aspecto relevante para la comparación, mientras que los parámetros de cambio de escala y la pendiente de las trayectorias son irrelevantes. Todos los métodos de reconocimiento de patrones utilizan la distancia entre objetos y prototipos de clústeres como un criterio de agrupamiento para determinar el grado de membresía de los objetos a los clústeres, mientras que la localización de los centros de los clústeres son obtenidos con base en la localización de los objetos en el espacio característico ponderado mediante su grado de membresía.

AGRADECIMIENTOS

Este archivo fue adaptado para informar la aceptación, la cual fue seleccionada por el Comité Académico del VII Encuentro de Investigación sobre Tecnologías de Información aplicadas a la Solución de Problemas EITI-2006.

REFERENCIAS

- Bakshi, B., Locher, G. y Stephanopoulos, G. (1994). 'Analysis of operating data for evaluation, diagnosis and control of batch operation', *Journal of process control* 4, 179-194. Butterworth-Heinemann.
- Ruger, R. (1989). *Induktive Statistik. Einführung für wirtschafts- und Sozialwissenschaftler*. R Oldenbourg Verlag, München, Wien.
- Setnes, M. y Kaymak, U. (1998). Extended fuzzy c-means with volume prototypes and cluster merging, in 'Proceedings of the 6th European Conference on Intelligent Techniques and Soft Computing (EUFIT'98). Aachen, Germany', pp. 1360-1364.