ISSN 0120-4211 BISTUA Vol. 5 No.1 Pag. 35 - 42

Similaridad Estructural de Sistemas Dinámicos Molecular con Base a sus Trayectorias

Castañeda Marín H [2], Rodríguez Graterol W[1], Colina Morles E [1], Chacón A.J [2]

[1]Universidad de los Andes, Mérida - Venezuela
[2]Facultad de Ciencias Básicas, Departamento de Física, Grupo Electrodinámica,Universidad de Pamplona, hcastaneda@unipamplona.edu.co

Recibido 14 Febrero 2007
Aceptado 09 Abril 2007

ABSTRACT

In Molecular Dynamic (MD) successive configurations are generated by integrating Newton's law of motion; the resulting trajectory specifies how the position and velocities of the particles in the system move with time.

The most cost consuming part is the computational calculation of forces on each particle from current positions, based on the force field. Generally MD uses simple models, where all collisions and are perfectly elastic; this occurs when the separation between the centers of the particles are equal to the point of discontinuity in the potential. When using continuous potentials, the force on each particle will change whenever the particle changes its position or whenever any of the other particles with which it interacts change position. The motions of all particles are coupled together, giving rise to a many body problem that cannot be solved analytically; finite different methods have to be used. The goals are to build an automated system to capture important events, such as, defect disintegration and defect amalgamation are detected using a dynamic fuzzy pattern recognition. The technique task of clustering methods is to partition a number of objects into small numbers of homogeneous clusters so that objects belonging to any one of the clusters would be as similar as possible and the object of different clusters as dissimilar as possible. The most important problem arising in this context is the choice of a relevant similarity measure, which is then used for definition of the clustering criterion.

KEY WORDS

Molecular Dynamic, Dynamic Fuzzy Pattern Recognition

RESUMEN

En la Dinámica Molecular (DM) una configuración sucesiva es generada mediante la integración de las leyes de movimiento de Newton, las trayectorias resultantes nos dan información acerca de como las posiciones y velocidades de las partículas en el sistema cambian con el transcurso del tiempo, en este contexto lo que mayor costo computacional exige es la determinación de las fuerzas aplicadas a cada partícula en su respectiva posición. Generalmente en la DM se suelen utilizar modelos simples, donde todas las colisiones son elásticas y ocurren cuando las separaciones entre los centros de las partículas son iguales al punto de discontinuidad del potencial.

Al utilizar potenciales continuos la fuerza y la posición de las partículas dependen de la interacción con las restantes partículas del sistema generando una interaccion de muchos cuerpos, por lo que no es posible solucionar analíticamente tal problema lo cual implica el uso de diferencias finitas.

La principal tarea en el método propuesto es particionar un número de objetos dinámicos en un pequeño numero de clústeres, de tal forma que los objetos en cada cluster sean en lo más posible similares y los objetos en diferentes clústeres son lo menos similares. El problema más importante en este contexto es la selección de una medida de similaridad pertinente, la cual será utilizada como criterio de agrupamiento.

El objetivo mas amplio del trabajo consiste en el análisis de un sistema molecular donde buscamos capturar importantes eventos por ejemplo la desintegración y fusión de defectos, la técnica utilizada es el reconocimiento de patrones temporales difusos. El problema más importante que se presenta en este contexto es la opción de una medida relevante de la similaridad, que se utiliza para la definición del criterio de agrupamiento.

PALABRAS CLAVES

Dinámica molecular, Reconocimiento de Patrones Dinámicos Difusos.

INTRODUCCIÒN

El interés principal de nuestra actividad investigativa es el descubrimiento de patrones dinámicos e información de un sistema estudiado, lo que puede redundar en un mayor compresión de la física que se esconde tras la evolución de los defectos estructurales en el estado solidó, su dinámica, todo esto en tiempo real. Lo anterior puede ser resultados de cambios abruptos en las estructura de los clústeres correspondiente a cambios de estado o comportamiento del sistema en consideración y que hace referencia a un cambio estructural. Debido al surgimiento de nuevos clústeres y la consideración de algunos datos históricos como irrelevantes, pueden aparecer cambios en la estructura dinámica del cluster [Man, 1983]

Si los nuevos datos no aparecen claramente asignados en los clústeres existentes, es necesario crear uno o más clústeres en forma secuencial o en paralelo. Esta situación puede aparecer si el grado de membresía del nuevo cluster con respecto a todos los clústeres difusos es igual o menor al de los ya formados.

Mezcla de clústeres: Dos o más clústeres pueden ser mezclados en un cluster si un gran número de datos tiene igualmente mayor grado de membresía (> 0.5) con respecto a los dos o mas clústeres.

División de clústeres: Un cluster puede convertirse en dos o mas clústeres, si un número grande de nuevos datos ha sido absorbido, distintos grupos de objetos con alta densidad dentro de un cluster puede ser formados, sin embargo el centro de un cluster puede ser localizado en el área de muy baja densidad.

En muchas áreas de la ciencia y la ingeniería, los sistemas pueden ser estudiados a través de la evolución de los rasgos temporales de sus propiedades observables. Diferentes tipos de sistemas o diferentes estados de un sencillo sistema puede distinguirse mediante un apropiado análisis de las secuencias temporales.

Usualmente la clasificación de series de tiempo es realizada mediante la computación de algunas características de los parámetros para cada serie de tiempo en cuestión. Las clasificaciones son realizadas con base a este conjunto de parámetros. Frecuentemente se hace en forma no supervisada, aunque no se conozca a priori que medición del parámetro corresponda a una determinada clase. En el caso de un rasgo escalar γ que sigue una distribución encorvada para distinguir k diferentes clases. La similaridad entre trayectorias puede ser interpretada en forma diferente dependiendo del contexto. En el lenguaje natural, la interpretación de similaridad esta asociada con "tener características comunes" o "no tener diferencia en la forma pero sí en tamaño o posición" [Setnes et el, 1998, p 378].

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.

Buscamos mostrar como el uso del conocimiento sobre la similaridad en series de tiempo, divide las secuencias en grupos significativos o clústeres.

Un conjunto mutuo de similaridades permite trabajar en un espacio abstracto de propiedades dinámicas sin tener que especificar una base o incluso una dimensión. Una aproximación similar produce resultados promisorios en el contexto especial de la clasificación de la morfología de los registros.

Fundamentos de Dinámica Molecular.

La dinámica molecular DM hace uso de las componentes del sistema en un sencillo esquema formado por: la energía potencial, la ecuación dinámica de cada una de las partículas, de donde se obtiene la aceleración y la velocidad para obtener finalmente las coordenadas y por consiguiente las trayectorias integrando las ecuaciones de movimiento de Newton a través de métodos numéricos, esperar generar trayectorias exactas sobre intervalos largos de tiempo cuando se ha integrado numéricamente con pasos de tiempo finito que no es factible, sin embargo esta exactitud no es necesaria, lo más importante es el comportamiento estadístico de la trayectoria para asegurar que las propiedades termodinámicas y dinámicas del sistema estén siendo obtenidas con una preedición suficiente, lo cual se cumple si el propagador del movimiento tiene la propiedad de simpleticidad o sea que conserva la métrica invariante del espacio de fases, lo cual implica a la vez que el error asociado al propagador esta acotado

$$\lim_{n_{paso}\to\infty}\left(\frac{1}{n_{paso}}\right)\sum_{k=1}^{n_{paso}}\left|\frac{\varepsilon(K\delta T)-\varepsilon(0)}{\varepsilon(0)}\right|\leq\varepsilon_{MD} \qquad (1)$$

Donde n_{paso} es el número de pasos de la simulación, $\varepsilon(0)=\mathrm{H}\left(r^N,p^N;0\right)$ la energía total inicial del sistema equilibrado, y ε_{MD} el límite superior que asegura la conservación de la energía Vg. 10^{-4} donde un valor de es aceptable, para sistemas Hamiltonianos, la propiedad de simpleticidad implica que el Jacobiano:

$$J\left(\Gamma_{\delta t},\Gamma_0\right)=\frac{dp(\Gamma^1_{\delta t},....,\Gamma^N_{\delta t})}{dp(\Gamma^1_0,...,\Gamma^N_0)} \qquad (2)$$

Sea unitario. Γ_0 Representa el vector inicial del espacio de fases de N dimensiones, que contiene todas las variables de posición, r, y

de impulsión, p , que describen el sistema. El potencial de Lenard –Jones es comúnmente utilizado para describrir la interaccion de sistemas compuestos por gases noble s como neon, Argón y por líquidos.

Al asumir un sistema conformado por átomos de argon tendremos básicamente primeros términos, de corto alcance y repulsivos, dado el principio de exclusión de Pauli, dos electrones no pueden ocupar la misma posición, lo cual termina manifestandose como una fuerza de repulsión entre cargas del mismo signo.

Atracción de largo alcance, segundo termino, los electrones alrededor del núcleo polarizado crean una atracción electrostática entre los átomos, par el caso átomos de Argon m=6.6x10 exp. -23 gramos, E=1.66x10 exp. -14 erg, sigma=3.4x 10 exp. 10-8

$$\vec{r}_1(t+\Delta) = 2\vec{r}_1(t) - \vec{r}_1(t-\Delta) + \vec{a}(t)\Delta^2 + O(\Delta^4)$$

$$\vec{v}_1(t) = \frac{\vec{r}_1(t+\Delta) - \vec{r}_1(t-\Delta)}{2\Delta} + O(\Delta^2) \qquad (3)$$

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. Considérese la trayectoria mostrada en la figura 1. Los patrones temporales en la trayectoria se pueden descomponer en segmentos indicando las tendencias locales; de tal forma que cada segmento esta limitado mediante puntos de inflexión o por un punto de inflexión y un punto extremo t , [Bakshi, et al., 1994]

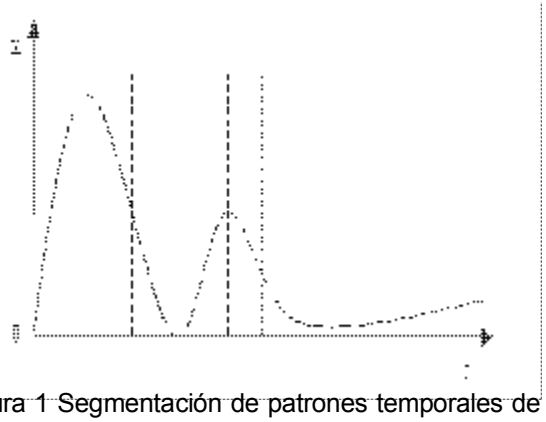


Figura 1 Segmentación de patrones temporales de una trayectoria de acuerdo a tendencias básicas.

En una trayectoria pueden ser distinguidos siete tipos de segmentos (tendencias), cada una de los cuales esta caracterizado mediante un signo constante en la primera y segunda derivada.

Dicha representación con tendencias triangulares proporciona unas características cualitativas para una descripción de los segmentos. Para derivar información cuantitativa a partir de los segmentos, estos son descritos mediante el siguiente conjunto

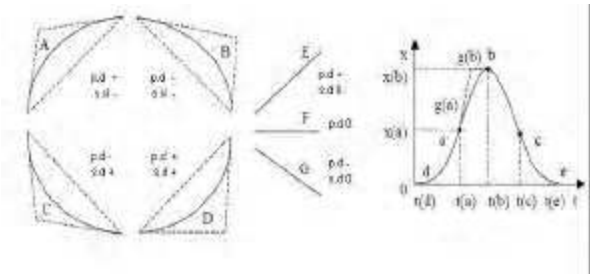


Figura 2 Características temporales cualitativas y cuantitativas obtenidas por segmentación.

$t(a)$, $t(b)$ Son los instantes de tiempo inicial y final del segmento. Por ejemplo, los instantes de tiempo de los valores de los puntos de inflexión y extremos locales de la trayectoria.

$x(a)$, $x(b)$ Son los valores de los puntos de inflexión y extremos locales de la trayectoria

$g(a)$, $g(b)$ Son los valores de la pendiente inicial y final del segmento. 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:

$gg(b)$ 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.

Md 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 limite de la trayectoria. Medidas estadísticas (ej. Valores de media, desviación estándar, coeficiente de correlación entre segmentos de trayectorias).

Definición 1. 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. Los parámetros listados anteriormente permiten una descripción precisa de la forma de los patrones temporales presentes en una trayectoria. Se toma en cuanta el número y tamaño de las colinas, sus pendientes, curvaturas y los instantes de sus ocurrencias y sus duraciones. Los factores de escala y traslación tienen efecto en los valores de los parámetros. Estas medidas de similaridad son adecuadas para el reconocimiento y comparación de patrones específicos en las trayectorias.

Definición del Problema .

La parte problema, se inicia con la obtención de las series de tiempo como fuentes que nos permite a partir de datos llegar a formar base de datos de conocimiento descubiertos (KDD).

Con un programa computacional con entradas en sus coordenadas y las velocidades de todos sus átomos. Los átomos son inicialmente arreglados en forma de un lattice regular. Estos átomos ocupan todas las esquinas y el centro de un cubo llamado celda unitaria. (Ver figura 5)

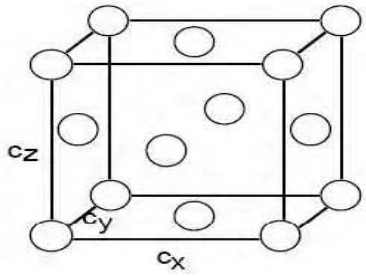


Figura 5 Lattice FCC

Aunque las celdas unitarias son cúbicas, $c_x = c_y = c_z = c$, cada celda contiene $\frac{1}{8}\text{x}8(esquinas)x\frac{1}{2}x6(caras) = 4atomos$ y la densidad numérica de datos esta dado por $\rho = \frac{4}{c^3}$ o $c = \left(\frac{4}{\rho}\right)^{\frac{1}{3}}$. Las 4 coordenadas están dadas como: $(0,0,0)$ $(0,1/2,1/2)$, $(1/2,0,1/2)$, $(1/2,1/2,0)$ estas cuatro ordenadas se almacenan en una matriz de orden 4x3.

El sistema simulado es construido mediante

la repetición del número de celdas unitarias en las direcciones x, y, z y almacenado cada una en un vector. El numero total de átomos esta dado entonces por:

$$N = 4xN_x^c xN_y^c xN_z^c \quad (4)$$

Se generan velocidades aleatorias de la magnitud $\frac{v_0^3}{3} = T_{inic}$, donde T_{inic} es la temperatura inicial así que:

$$v_0 = \sqrt{3T_{inic}} \quad (5)$$

Para cada átomo el vector de velocidad esta dado mediante:

$\vec{v}_i = v_0(\zeta_0, \zeta_1, \zeta_2) = \vec{v}_o$ Donde $\vec{\zeta}$ es un vector orientado aleatoriamente de longitud unitaria.

$$\vec{v}_i = v_0(\zeta_0, \zeta_1, \zeta_2) = \vec{v}_o \quad (6)$$

La secuencia de enteros aleatorios aparentemente entre 0 y $m-1$ es obtenida mediante una inicialización para I_1 diferente de cero. La ecuación recursiva dada por $I_{j+1} = I_j \pmod m$, donde se selecciona $m = 2^{31} - 1 = 2147483647$ y $a = 16807$, obteniendo los números aleatorios uniformes en el rango $[0,1]$ mediante:

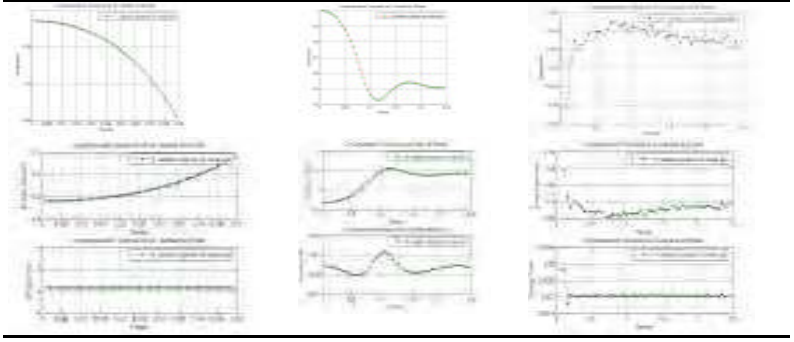
$$r_j = I_j / m \quad (7)$$

RESULTADOS

A continuación se ilustra el comportamiento de un sistema dinámico molecular específicamente a través de Molecular dynamics (MD) simulation with the Lennard-Jones potential. donde se observa el comportamiento de las variables de estado, temperatura , energía potencial y energía total.Se quiere analizar los rasgos del comportamiento de las variables de estado en términos de similaridad estructural entre dos trayectorias

La definición de similaridad estructural se ilustrada 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 similaridad $s(\mathbf{x},\mathbf{y})$, donde el parámetro a_1 se usa como característica K de la trayectoria y el conjunto difuso A denota “admisible iferencia para la tendencia” , esta similaridad estructural puede

Tabla No.1 Trayectorias de comportamiento de la Temperatura, Energía potencial y Total con 3, 6,11 celdas unitarias por cada eje.

Variable Estado (3 cu)	Variable estado (6cu)	Variable estado (11cu)
		

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 similaridad , donde la característica es un vector de coeficientes de la segunda derivada el conjunto difuso 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 similaridad se define para cada combinación ; en esta forma la máxima similaridad corresponde al mejor emparejamiento de las trayectorias con respecto a la curvatura encontrada. La medida de similaridad 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

Los parámetros listados en la similaridad estructural basada en los parámetros temporales específicos de las trayectorias permiten una descripción de la forma de patrones temporales presentes en las trayectorias.;ellos tienen en

Tabla No.2 Ilustra la similaridad estructural entre el comportamiento temporal de la temperatura para 3 y 6 celdas unitarias y entre 6 y 11 celdas unitarias

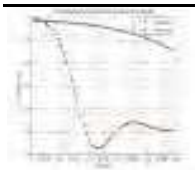
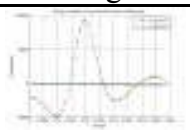
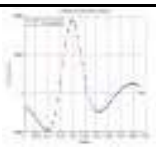
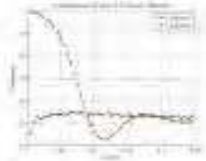
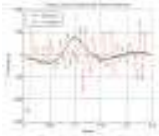
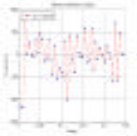
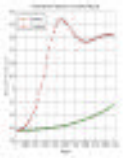
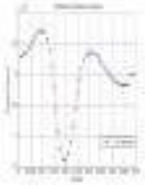
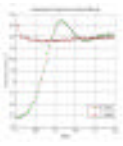
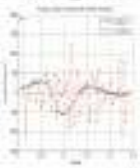
variable	Rasgo	Dif	C.difuso
			
			

Tabla No.3 Ilustra la similaridad estructural entre el comportamiento temporal de la Energía potencial para 3 y 6 celdas unitarias y entre 6 y 11 celdas unitarias.

Variable	Rasgo	Dif erencia	C. F
			
			

cuenta el número y tamaño de las colinas, su pendiente y curvatura, el instante de su apariencia y su duración, donde los factores de cambio de escala y traslación tienen un efecto en los valores de los parámetros, esta medida de similaridad está disponible para el reconocimiento y comparación de patrones específicos de trayectorias

CONCLUSIONES

Las anteriores definiciones sobre medidas de similaridad estructural pueden ser utilizadas en diferentes combinaciones para obtener una evaluación más completa de la similaridad entre trayectorias. En algunos casos la similaridad estructural puede ser reducida a similaridad puntual. Por ejemplo, considerando la similaridad 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 similaridad 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.

REFERENCIAS

Bakshi, B., R., Locher, G., Stephanopoulos, G., (1994), Analysis of operating data for Evaluation, diagnosis and control of Batch operation. Journal of process control, Vol 4, 1994, Butterworth-Heinemman, 179-194.

Das, G., Gunopulos, D., Mannila, H. (1997), Finding Similar Times Series. In: Komorowski, J., Zytkow, J. (Eds) Principles of Data Mining and Knowledge Discovery. Proceedings of the First Europeam Symposium PKDD´97, Trondheim, Norway 1997, Springer, 1997 .pp 80-100.

Joentgen., Mikenina. Weber, R., Zimmerman, H.-J., (1998), Dynamic Fuzzy Data Analysis: Similarity between Trajectories. In: Bauer. (Ed.) Fuzzy Neuro System' 98, computational intelligency, Sankt, augusting, p 98-105.

Nemirko, A.P., Manilo, LA., Kalinichenko, A.N. (1995) Waveform Classification for Dynamic Analysis of ECG. Pattern recognition and Image, Analysis, Vol. 5 (1), 1995, p. 131-134

Pedrycz, W. (1990) Fuzzy Sets in Pattern Recognition: Methodology and Methods. Pattern Recognition, Vol. 23,1990, p. 121-146

Pedrycz, W. (1990) Fuzzy Sets in Pattern Recognition: Accomplishments and Challenges. Fuzzy Sets and Systems, 90, 1997, p. 171-176

Ruger. (1989) Induktive Statisk, Einführung für wirtschafts –und Sozialwissenschaftler. R Oldenbourg Verlag, Müchen, Wien, 1989

Schreiber, T., Schmitz, A. (1997) Classification of Time Series Data with Nonlinear Similarity Measures. Physical Review Etters, Vol. 79 (8), 1997, p. 1475-1478

Setnes, M., Kaymak, U. (1998) Extended Fuzzy c-Means with Volume Prototypes and Cluster Merging. Proceedings of the 6th European Conference on Intelligent Techniques and Soft Computing (EUFIT'98), Aachen, Germany, September 7-10, 1998, p. 1360-1364

Taylor, C. Nakhaeizadeh, G., Lanquillon, C. (1997) Structural Change and Classification. In G. Nakhaeizadeh, I. Bruha, C. Taylor (Eds.) Workshop Notes on Dynamically Changing Domains: Theory Revision and Context Dependence Issues, 9th European Conference on Machine Learning (ECML'97), Prague, Czech Republic, p. 67-78