

Clasificación de Trayectorias en Dinámica Molecular Usando Relaciones de Equivalencia Difusa y Análisis de Componentes Principales

Classification of Dynamic Molecular Trajectories using Fuzzy Equivalence Relation and Principal Component Analysis

Hernando Castañeda M. Msc, Wladimir Rodríguez G., PhD, y Eliézer Colina M. PhD
 Universidad de Pamplona, Doctorado en Ciencias Aplicadas ULA, Facultad de Ingeniería, Departamento de Computación ULA, Facultad de Ingeniería, Departamento de Automatización y Control
hcstaneda@ula.ve, wladimir@ula.ve, ecolina@ula.ve

Recibido para revisión 26 de Marzo de 2007, aceptado 15 de Junio de 2007, versión final 20 de junio de 2007

Resumen—En 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 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 interacción de muchos cuerpos, una solución analítica no es posible aun en este problema. Existe una aproximación de potenciales atómicos que interactúan con funciones de entrenamiento analíticas mediante puntos de Datos.

La principal tarea en el método propuesto es generar un conjunto de datos de entrenamiento mediante la acumulación de la funciones de potencial de Lennard Jones, extraer los rasgos estructurales a partir de sus trayectorias y particionar un número de objetos dinámicos en un pequeño número 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, permitiendo predecir el comportamiento de las variables de salida.

El objetivo más 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 concepto 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 Clave— Dinámica de Sistemas, Lógica Difusa, Aplicaciones de Inteligencia Artificial

Abstract—In Molecular Dynamic, 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 expensive part is the calculation of forces on each particles from current positions, based on the force field, when MD using simples models all collisions are perfectly elastic and occur when the separation between the centers of the particles equal to point of discontinuity in the potential. When is using with continuous potentials, the force on each particle will change whenever the particle change 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 difference methods has to be used.

The main task in the proposed method is to generate a group of data of training by means of the accumulation of the functions of potential of Lennard Jones, to extract the structural features starting from its trajectories and a number of dynamic objects segment in a small clusters number, in such a way that the objects in each cluster are in the most possible thing similar and the objects in different clusters are the less similar ones, allowing to predict the behavior of the exit variables

The goals is build an automated system to capture important events such as defect disintegration and defect amalgamation but goals initial is to understand the interaction among defect using

dynamic fuzzy pattern recognition . The task of clustering methods is to partition a number of objects into small numbers of homogeneous clusters so that objects belonging to any one the clusters would be as similar as possible and object of different clusters as dissimilar possible. The most important problem arising in this context is the choice of a relevant similarity measure, which is use for definition of the clustering criterion

Key words—Dynamic System, Fuzzy Logic, Application of Intelligent System

INTRODUCTION

El interés principal de nuestra actividad investigativa es el Edescubrimiento de patrones dinámicos e información de un sistema estudiado, lo que puede redundar en un mayor comprensión de la física que se esconde tras la evolución de los defectos estructurales , su dinámica, esto en tiempo real. Lo anterior pueden 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]

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 en diferentes estados de un sistema sencillo 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 los nuevos datos que 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.

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 similitud entre trayectorias puede ser interpretada en forma diferente dependiendo del contexto. En el lenguaje natural, la interpretación de similitud esta asociada con “tener características comunes” o “no tener diferencia en la forma pero sí en tamaño o posición” [Setnes et al, 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 similitud 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 similitud en series de tiempo, divide las secuencias en grupos significativos o clústeres.

Un conjunto mutuo de similitudes 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.

I. FUNDAMENTOS DE DINÁMICA MOLECULAR

La dinámica molecular 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, 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 mediante de métodos numéricos; se espera generar trayectorias exactas sobre intervalos largos de tiempo ; cuando se ha integrado numéricamente con pasos de tiempo finito 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 simplecticidad 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

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

Donde n_{paso} es el número de pasos de la simulación, $\varepsilon_0 = H(r^n, p^n, o)$ es la energía total inicial del sistema equilibrado, y ε_{MD} es 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 simplecticidad implica que el Jacobiano sea unitario:

$$J(\Gamma_{\delta t}, \Gamma_0) = \frac{\partial (\Gamma_{\delta t}^1, \dots, \Gamma_{\delta t}^N)}{\partial p(\Gamma_0^1, \dots, \Gamma_0^N)} \quad (2)$$

Donde Γ_0 Representa el vector inicial del espacio de fases de N dimensiones, contiene todas las variables de posición r y de impulsión p que describen el sistema. El potencial de Lenard -Jones es comúnmente utilizado para describir la

interacción de sistemas compuestos por gases nobles como Neón, Argón y para líquidos.

Al asumir un sistema conformado por átomos de argon se tiene básicamente que los primeros términos (ecuación 3) son de corto alcance y repulsivos y según el principio de exclusión de Pauli, dos electrones no pueden ocupar la misma posición, lo cual termina manifestándose como una fuerza de repulsión entre cargas del mismo signo.

La atracción de largo alcance, segundo termino (ecuación 3), los electrones alrededor del núcleo polarizado crean una atracción electrostática entre los átomos, par el caso átomos de Argon $m=6.6 \times 10^{-23}$ gramos, $E=1.66 \times 10^{-14}$ erg, $\sigma=3.4 \times 10^{-8}$

$$U(r) = \varepsilon \left[\left(\frac{b}{r} \right)^6 - 2 \left(\frac{b}{r} \right)^{12} \right] \quad (3)$$

II. NOCIONES DE SIMILARIDAD ESTRUCTURAL PARA TRAYECTORIAS

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, (referencia). 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.

A. Similaridad en base a Curvatura de las Trayectorias

La curvatura de las trayectorias de cada punto describe el grado con la cual una trayectoria esta 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).

$$cv_k = \ddot{x}_k = \frac{\dot{x}_k - \dot{x}_{k-1}}{t_k - t_{k-1}} \quad k = 3, \dots, p \quad (4)$$

Donde \dot{x}_k denota el coeficiente de la primera derivada en el punto x_k dado por:

$$\dot{x}_k = \frac{x_k - x_{k-1}}{t_k - t_{k-1}} \quad k=2, \dots, p \quad (5)$$

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

$$cv_k = \ddot{x}_k = \frac{x_k - 2x_{k-1} + x_{k-2}}{(t_k - t_{k-1})^2} \quad k = 3, \dots, p \quad (6)$$

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 Si 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 trayectorias $x(t)$ e $y(t)$ son consideradas similares con respecto a su curvatura si están caracterizadas por coeficientes similares en su segunda derivada.

Esta definición de similaridad estructural es ilustrada en la figura 1, donde tres tipos de trayectorias están representadas: la trayectoria convexa C, la trayectoria cóncava B, y la trayectoria A con comportamiento oscilatorio que cambia de cóncavo a convexo.

La medida de la similaridad basada en la curvatura es particularmente adecuada para trayectorias con un bajo número de fluctuaciones y en forma ondulada. Esta medida sin embargo es susceptible al cambio de escala, por ejemplo una trayectoria transformada mediante un factor de escala tiene una curvatura diferente que al comienzo.

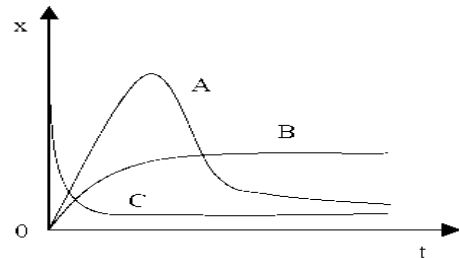


Figura 1. Similaridad Estructural basada en Curvatura de las Trayectorias.

B. Agrupamiento por relación de equivalencia difusa para la identificación de los rasgos.

Como los rasgos de las trayectorias que caracterizan el comportamiento de un sistema dinámico no pueden definirse crisp, pero puede ser aproximado como un subconjunto difuso en un espacio bidimensional definido por la curvatura y la pendiente. Sin embargo la aproximación utilizada para la clasificación esta basada en el concepto de equivalencia difusa, que define una relación que es reflexiva, simétrica y transitiva máx.-min. Este algoritmo se ilustra mediante el diagrama de flujo que se muestra en la figura 2. Una relación de equivalencia difusa define una partición difusa en cada nivel α , cada una de las cuales representa un cluster separado.

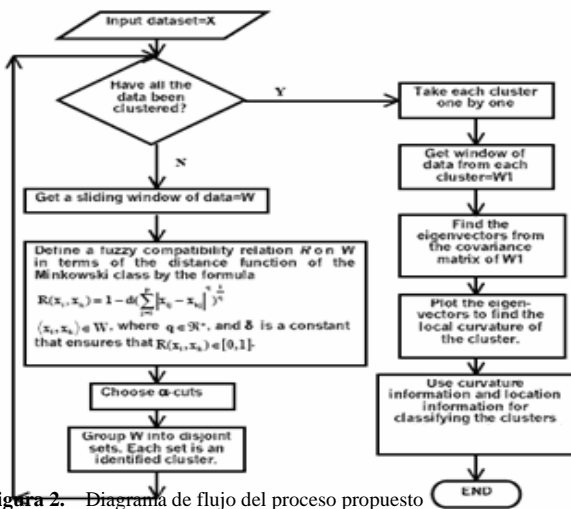


Figura 2. Diagrama de flujo del proceso propuesto

El análisis de componentes principales (PCA) es usado para encontrar la curvatura del cluster identificado, aunque los datos exhiben una variación considerable, la cual no puede estar representada exactamente mediante los primeros pocos vectores propios del conjunto completo de datos, sin embargo PCA es ejecutada en una ventana de datos, para encontrar las orientaciones locales.

III. CLUSTERING MEDIANTE RELACIONES DE EQUIVALENCIA DIFUSA

Esta sección describe una técnica clustering en la cual los datos multivariados son usados para formar una matriz de relación de equivalencia difusa. Diferentes niveles α de este conjunto difuso se producen en número diferentes de clústeres a partir de los datos originales. Dado un conjunto de datos $\{(x_{11} \dots x_{1p}) \dots (x_{n1} \dots x_{np})\}$ con n muestras sobre un espacio p -dimensional de rasgos P , una matriz de relaciones de equivalencia difusa Q con dimensiones $n \times n$ es determinada. Defínase M_k y m_k el valor máximo y mínimo respectivamente de x_{jk} para cada rasgo k en P como:

$$M_k = \frac{\max}{j \in \{1, 2, \dots, n\}} x_{jk} \quad Y \quad (7)$$

$$m_k = \frac{\min}{j \in \{1, 2, \dots, n\}} x_{jk}$$

Definiendo el término i, j enésimo de la matriz Q como:

$$Q_{ij} = 1 - \frac{1}{p} \sum_{k=1}^p \left(\frac{|x_{ik} - x_{jk}|}{M_k - m_k} \right)^s \quad (8)$$

Para formar la matriz relacional de compatibilidad difusa Q . Cada q_{ij} representa la composición de la relación Q de la forma $Q \circ Q$. La clausura T , de Q puede ser computada mediante el siguiente procedimiento.

Haga

$$T' = Q$$

$$Q = T' \circ T'$$

hasta $Q(Q \neq T')$

$$T = T'$$

Desde la matriz de clausura transitiva T con elementos t_{ij} , una colección de clústeres C , es formado para un grado de membresía específico α . Sea $C_k \in C$ tal que $\forall i, j \in C_k, t_{ij} \geq \alpha$ forman una clase de equivalencia difusa.

Se define un cluster de equivalencia difusa W_i mediante $\{(x_{w1}, \dots, x_{wp}) | w \in c_i\}$.

1.1 Conjuntos de Múltiples clusters.

Definamos el centroide del cluster W_i como:

$$a_i = \frac{1}{|C_i|} \sum_{j \in C_i} x_j \quad (9)$$

Donde $|C_i|$ denota la cardinalidad del conjunto de clústeres C_i . Sea $A_\alpha = \{a_1, \dots, a_m\}$ el conjunto de centroides de clusters difusos resultado de coleccionar datos durante un intervalo de tiempo, Δt_{k1} y $B_\alpha = \{B_1, \dots, B_m\}$ un segundo conjunto de centroides de clusters difusos resultado de coleccionar datos durante un intervalo Δt_{k2} donde $k2 > k1$ con máximos y mínimos acumulativos para cada

rasgo $\max_{A_k}, \max_{B_k}, \min_{A_k}, \min_{B_k}$ donde se obtienen los términos sin perdida de generalidad:

$$\max_{A_k} = \frac{\max_i \max_{j \in C_i} x_{jk}}{i} \quad \text{Y} \quad \min_{A_k} = \frac{\min_i \min_{j \in C_i} x_{jk}}{i} \quad \text{los}$$

vectores de peso $w_{A\alpha} = \{w_{a1}, \dots, w_{am}\}$ y $w_{B\alpha} = \{w_{b1}, \dots, w_{bm}\}$ formalizados sin perdida de generalidad mediante $W_{\alpha\alpha} = \{w_{ai} | w_{ai} = |C_i|\}$.

Defínase el rango global como: $r_k = \max(\max_{A_k}, \max_{B_k}) - \min(\min_{A_k}, \min_{B_k})$, la matriz relacional difusa será de la forma:

$$Z_{ij} = 1 - \frac{1}{p} \sum_{k=1}^p \left(\left| \frac{a_{jk} - b_{ik}}{r_k} \right|^s \right)^{1/s} \quad (10)$$

Donde p es la dimensión del espacio de rasgos y $s = 1$ para la distancia de Hamming y $s = 2$ para la distancia Euclidiana.

Con las proyecciones $\rho_i^A = \frac{\max_j z_{ij}}{i}$ y $\rho_i^B = \frac{\max_j z_{ij}}{j}$

1.2 Conjunto de Mezcla de Clusters.

Usando la matriz de relación centroide Z y ρ_i^B , una nueva colección de clústeres N_{t+1} es construida con un umbral β que satisface $\forall \rho_i^B > \beta, a_j$ es remplazado así:

$$\frac{w_{aj}a_j + w_{bi}b_i}{w_{aj} + w_{bi}} \Big|_{z_{ij}} = \rho_i^B \quad (13)$$

Donde en consecuencia w_{aj} es remplazado por $w_{bi} + w_{aj}$. Finalmente la forma:

$$N_{t+1} = \left\{ b_i \mid \rho_i^B \leq \beta \right\} \cup A_\alpha \quad (11)$$

Para una nueva colección de clústeres representando series de tiempo Δt_{k1} y Δt_{k2} . Este método debe ser repetido para cada sucesivo intervalo de tiempo.

1.3 Medida de Similitud en el Cluster.

Sea A_α el conjunto de clústeres formados mediante adición o mezcla de conjuntos de clústeres en un número de intervalos de tiempo. Sea B_α el conjunto de clusters durante un nuevo intervalo de tiempo Δt_n . una matriz de relación difusa Z puede ser computada y la proyección β_i^B . La medida de similitud q es definida como el grado en la cual el conjunto

de clústeres B_α es similar al conjunto de clusters A_α y formalizada así:

$$q = \frac{1}{n} \sum_{i=1}^n \rho_i^B \quad (12)$$

Ejemplos del uso de la medida de similitud en sistemas dinámicos se dará en la siguiente sección:

IV. DEFINICION DEL PROBLEMA

La parte problema 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 (BCD).

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)

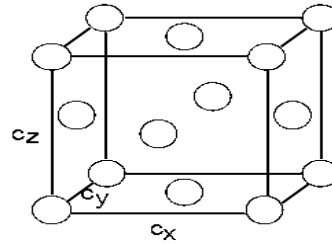


Figura 3. Lattice FCC.

Aunque las celdas unitarias son cúbicas, $c_x = c_y = c_z = c$, cada celda contiene $\frac{1}{8} \times 8(\text{esquinas}) + \frac{1}{2} \times 6(\text{caras}) = 4 \text{ átomos}$ 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 4×3 .

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 número total de átomos esta dado entonces por:

$$N = 4xN_x^c \times yN_y^c \times zN_z^c \quad (13)$$

Se generan velocidades aleatorias de la magnitud $\frac{v_0^3}{3} = T_{inic}$,

donde T_{inic} es la temperatura inicial así que:

$$v_0 = \sqrt{3T_{inic}} \tag{14}$$

Para cada átomo el vector de velocidad esta dado mediante:

$$\vec{v}_i = v_0 (\zeta_0, \zeta_1, \zeta_2) = \vec{v}_o \tag{15}$$

Donde $\vec{\zeta}$ es un vector orientado aleatoriamente de longitud unitaria.

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:

$$r_j = I_j / m \tag{16}$$

V. RESULTADOS

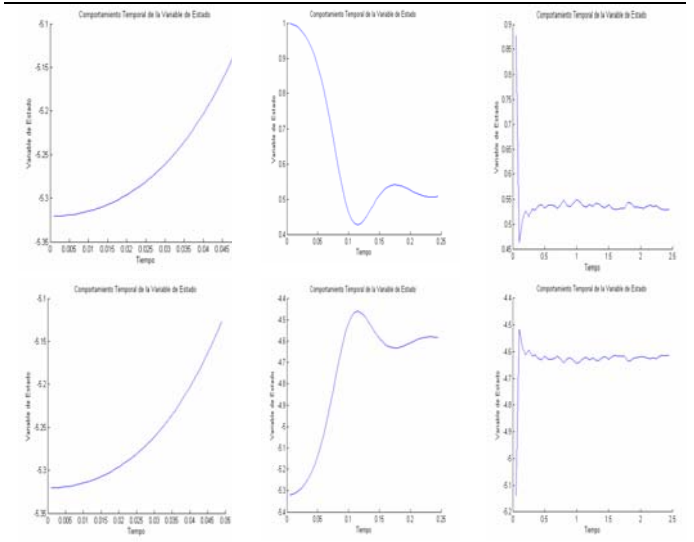
A continuación se ilustra el comportamiento de un sistema dinámico molecular específicamente a través de “Molecular dynamics (MD) simulation with the Lennard-Jones potencial” donde se observa el comportamiento de las variables de estado, temperatura, energía potencial y energía total. Se quiere analizar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

En la tabla 1 se ilustra el comportamiento temporal de la temperatura y la energía potencial con tamaños de 3,6, y 11 a celdas unitarias por cada eje. Estas trayectorias son la base para extraer puntualmente su pendiente y curvatura que se convierten en componentes temporales en el espacio de rasgos.

La definición de similaridad estructural donde las trayectorias presentan diferentes comportamientos a medida que se incrementa el número de celdas unitarias en sus ejes.

Tabla 1. Trayectorias de comportamiento de la Temperatura, Energía potencial con 3, 6,11 celdas unitarias por cada eje.

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



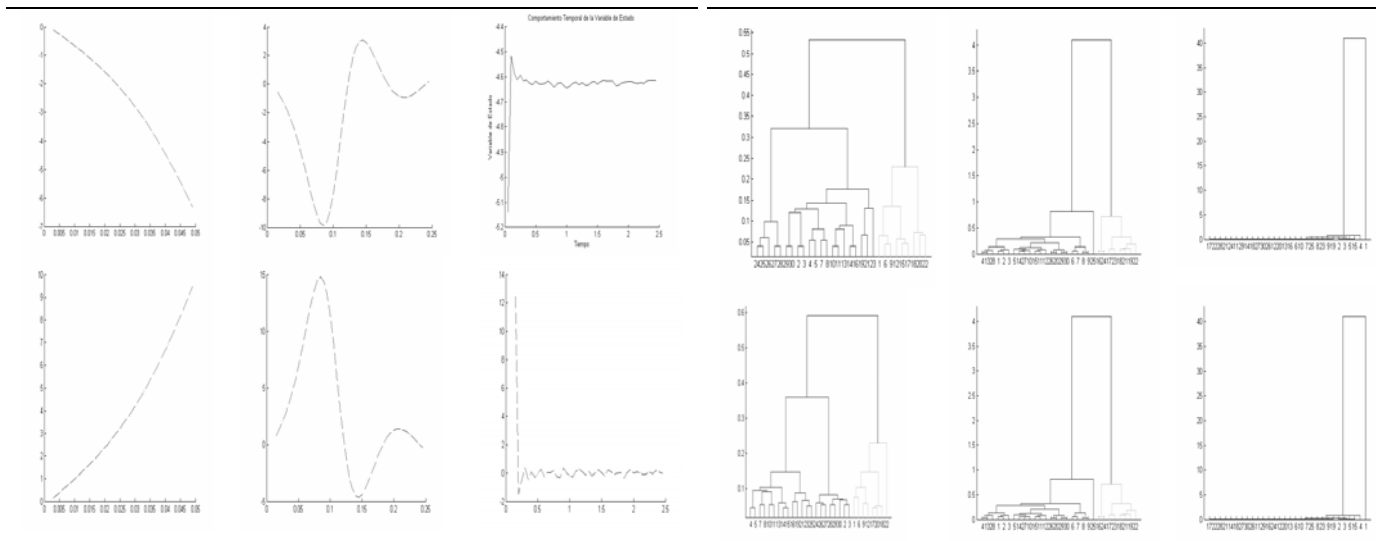
La medida de similaridad $s(\mathbf{x}, \mathbf{y})$, donde la característica K es un vector de coeficientes de la segunda derivada $K = [cv_3, cv_4, \dots, cv_p]$ el conjunto difuso A denota “diferencia admisible para la relación difusa entre pendiente y curvatura” si la traslación temporal es irrelevante 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.

En la tabla 2 se ilustra el comportamiento temporal para la temperatura y energía potencial en el espacio de rasgos (pendientes y curvaturas puntuales y temporales) para cada una de las tres celdas en los ejes en consideración.

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 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.

Tabla 2. Ilustra la similaridad estructural entre el comportamiento temporal de la temperatura y energía potencial para 3, 6,10 celdas unitarias.

IPC (3 cu)	IPC (6cu)	IPC (11cu)
------------	-----------	------------



Este trabajo muestra las trazas de las variables de estado (temperatura y energía potencial) que no pueden ser definidas robustamente pero que pueden aproximarse como subconjuntos difusos en un espacio bidimensional definido mediante la pendiente y la curvatura.

En la tabla 3 se ilustran los dendogramas (generados en el proceso de relacion de equivalencia difusa utilizando como metrica la distancia euclidiana) que ilustran el comportamiento de la temperatura y energía potencial para 3, 6, 11, celdas unitarias en los ejes a partir de un numero de niveles α (umbrales) que subdividen los datos en diferentes subconjuntos.

Una vez que los datos son agrupados, el análisis de componentes principales es aplicado para extraer estructuras de alta dimensión con el propósito de encontrar auto descomposiciones el espacio de patrones.

Este método es particularmente útil cuando en el espacio de patrones original puede ser descrito con precisión en un subespacio medido mediante mucho del primer auto vector.

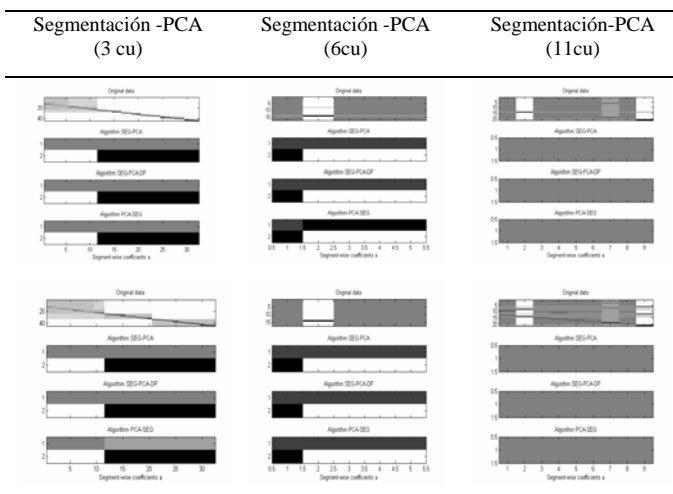
El propósito de la segmentación es la descomponer la secuencia en unos pequeños número de subconjuntos homogéneos llamados segmentos, de tal forma que los datos en cada segmento pueda ser descrito con precisión mediante un simple modelo, por ejemplo una constante mas ruido.

Los algoritmos de segmentación son ampliamente utilizados para extraer estructuras a partir de secuencias y por ello se utilizaron tres algoritmos el SEG-PCA, el SEG-PCA-DPA y el PCA-SEG [Ella Bingham et la, 2000]. En la tabla 4 se ilustra básicas segmentaciones para Cada uno de los clústeres obtenidos en el proceso de relacion de equivalencia difusa y representada mediante los dendogramas. En la tabla 4 se ilustra una muestra del proceso.

Tabla 3. Ilustra el agrupamiento para el comportamiento temporal de la temperatura y Energía potencial para celdas unitarias 3, 6,11.

Dendograma (3 cu)	Dendograma (6cu)	Dendograma (11cu)
-------------------	------------------	-------------------

Tabla 4. Segmentaciones básicas de la Temperatura y Energía potencial para celdas unitarias 3, 6,10.



VI. CONCLUSIONES

La figura 4 muestra los datos originales, acompañados de los coeficientes docto- segmentos como imágenes en escala de grises para cada uno de los tres algoritmos tipo PCA mencionados anteriormente, para segmentos de tamaño 2. se notan zonas (rasgo en la intensidad del color) de comportamiento similar en el comportamiento de las variables temperatura y energía potencial entre los diferentes números de celdas unitarias. También se puede observar como es de fuerte la estructura latente tanto en la temperatura como en la energía potencial en el caso de considerarse 11 celdas unitarias en los ejes. Asi que diferentes segmentos en algunos casos reales poseen básicamente fundamentos comunes.

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 de rasgos

ponderado mediante su grado de membresía. En el agrupamiento usando relaciones de equivalencia no es necesario especificar el número de clases.

RECONOCIMIENTO

Expresamos agradecimiento a la Universidad de Pamplona (Col) la oportunidad manifiesta de sus directivas encabezada por el señor rector Álvaro González Joves para que su recurso humano se cualifica a nivel de doctorado y la Universidad de los Andes (Ve) por ampliar sus conocimiento a través de su recursos muy bien preparados en diferentes instituciones educativas del mundo, a alumnos principalmente a diferentes países Latinoamericanos.

REFERENCIAS

- [1] 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.
- [2] 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 European Symposium PKDD'97, Trondheim, Norway 1997*, Springer, 1997 .pp 80-100.
- [3] 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.
- [4] 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
- [5] Pedrycz, W. (1990) *Fuzzy Sets in Pattern Recognition: Methodology and Methods. Pattern Recognition*, Vol. 23,1990, p. 121-146
- [6] Pedrycz, W. (1990) *Fuzzy Sets in Pattern Recognition: Accomplishments and Challenges. Fuzzy Sets and Systems*, 90, 1997, p. 171-176
- [7] Ruger. (1989) *Induktive Statistik, Einführung für wirtschafts –und Sozialwissenschaftler*. R Oldenbourg Verlag, München, Wien, 1989
- [8] Schreiber, T., Schmitz, A. (1997) *Classification of Time Series Data with Nonlinear Similarity Measures. Physical Review E*, Vol. 79 (8), 1997, p. 1475-1478
- [9] 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
- [10] 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
- [11] Ella Bingham , Aristides Gionis,Heli asila, Heikki Mannila, Evimaria Terzi, *segmentation and dimensionality reduction, HIIT BASIC RESEARCH UNIT , University of Helsinki and Helsinki university of technology Finland , 200*