

ESPECTROS DE EVIDENCIA TAXONOMICA EN BASES DE DATOS

Gregorio Perichinsky Elizabeth Jiménez Rey María Delia Grossi

{ gperi, ejimenez, mdgrossi }@mara.fi.uba.ar
Laboratorio de Bases de Datos y Sistemas Operativos
Departamento de Computación - Facultad de Ingeniería
Universidad de Buenos Aires
Paseo Colón N°850 - (1063) Buenos Aires - Argentina
Phone: (54 1) 343-1177 (int. 140/145)
FAX: (54 1) 331-0129

Resumen.

La Taxonomía Numérica permite agrupar, a través de métodos numéricos llamados análisis de clusters, unidades taxonómicas operacionales (OTUs) en taxa (grupos de taxones u OTUs). Los clusters constituyen familias mediante un análisis estructural basado en su característica fenotípica, en grado de similitud, entre dos OTUs o grupos de OTUs para clasificar formando familias o clusters.

Las OTUs toman valores de los dominios dinámicos de atributos que forman entidades y aplicando la semántica del modelo de datos de Bases de Datos Relacionales dinámicas se representan los objetos taxonómicos.

Las familias de OTUs, que se asocian por su grado de similitud, se obtienen mediante la distancia Euclídea y la aplicación de técnicas de "*vecinos próximos*", lográndose la fuente de la **evidencia taxonómica** al cuantificar, a partir del coeficiente de similitud de semejanza para cada par de OTUs de la matriz básica de datos.

Lo novedoso es el concepto de espectro de los estados de los caracteres de los pares de OTUs respecto al total, el espectro de familias, por el principio de superposición al procesar los espectros de los pares de OTUs y la obtención de invariantes (centroide, varianza y radio).

Finalmente se logra un algoritmo más eficiente por un mejor tratamiento matricial.

Introducción.

La clasificación es una técnica de abstracción usada para agrupar objetos con propiedades comunes. Esta permite delimitar el dominio de objetos, bajo la hipótesis de que cada objeto pertenece a una clase (y solo una) y de que, para cada clase, hay al menos un objeto que pertenece a ella.

Se buscan conceptos clasificatorios mediante nuevas herramientas analíticas que permitan una estructura de clasificación que no se modifique por el agregado de nueva información (estabilidad de la clasificación), ni se altere por la incorporación de nuevas entidades (Perichinsky, G. et al, Innsbruck. 1996-1997).

Este trabajo es un nuevo avance de la taxonomía numérica con Bases de Datos Relacionales Dinámicas, que constituyen un modelo que crea dominios independientes dinámicamente integrables por medio de estructuras dinámicas.

Algoritmo.

La secuencia algorítmica está dada por la conformación de la matriz de datos, su normalización, la construcción de la matriz de similitud, los espectros de objetos y de familias formadas por clustering y el análisis de invariantes.

Corolario.

El tratamiento dinámico e integrado de los dominios permite una fácil normalización, atributo - dominio - valor, y la implementación en el modelo de Base de Datos Dinámica y su utilización en Taxonomía Numérica.

La contribución teórica-empírica es la aglomeración de objetos formando clases producidas por pasos del método (algoritmo) obteniendo clusters y dominios con valores normalizados y la **densidad** y el **rango** en términos del **RADIO** del conjunto que es visualizado como una **INVARIANTE CARACTERÍSTICA de los OTUs.**