

Tabla de Contenido

1. Introducción	1
1.1. Motivación	1
1.2. Objetivos	2
1.2.1. Objetivo General	2
1.2.2. Objetivos Específicos	2
1.3. Alcances y Limitaciones	3
1.4. Estructura de la Tesis	3
2. Antecedentes y Marco Teórico	5
2.1. Introducción	5
2.2. Clustering	5
2.2.1. Distancias y Similitudes	6
2.2.2. Maldición de la Dimensionalidad	8
2.2.3. Tipos de Algoritmos	10
2.2.4. Similaridades entre Clusters	12
2.3. Clustering en Subespacios	13
2.4. Métodos de Resolución	16
2.5. Dinamismos en Clustering	16
2.5.1. Dinamismo dentro de los objetos	16
2.5.2. Dinamismo de nuevos objetos	19
2.5.3. Dinamismo en nuevos atributos	19
2.6. Descripción del Modelo Base (Proclus)	19
3. Descripción de la Metodología propuesta	23

3.1. Introducción	23
3.2. Metodología General Propuesta	23
3.3. Implementación de la Metodología General Propuesta	26
4. Implementación	29
4.1. Introducción	29
4.2. Simulación de datos	29
4.3. Simulación de datos	31
4.4. Resultados	34
4.5. Robustez de los resultados.	37
4.6. Convergencia de los Resultados	40
5. Conclusiones	42
5.1. Conclusiones Generales	42
5.2. Sobre la metodología y método implementado	43
5.3. Trabajos Futuros	43
Bibliografía	45
A. Algoritmos Capítulo 2	47
B. Modificación Algoritmo Capítulo 3	50
C. Resultados	52

Índice de Tablas

2.1. Definiciones de ejemplos de distancias	7
2.2. Propiedades de distancias	8
2.3. Distancias inter-cluster más comunes	13
2.4. Distancias intra-cluster más comunes	13
4.1. Dimensiones Datos 01	32
4.2. Dimensiones Datos 02-01	33
4.3. Dimensiones Datos 02-02	33
4.4. Dimensiones Datos 03-01	34
4.5. Dimensiones Datos 03-02	34
4.6. Comparación de tiempos de ejecución de los métodos	35
4.7. Puntos asignados a cada cluster por iteración (caso estático)	35
4.8. Comparación de tiempos de ejecución de los métodos	36
4.9. Puntos asignados a cada cluster caso 2 por iteración (caso dinámico)	36
4.10. Matriz de confusión para el caso 2 (caso dinámico)	36
4.11. Comparación de tiempos de ejecución de los métodos caso 03	37
4.12. Matriz de confusión para el caso 3 (caso dinámico)	37
4.13. Comparación de tiempos promedio de ejecución de los métodos.	38
C.1. Dimensiones relevantes detectadas Caso 2 parte 01	52
C.2. Dimensiones relevantes detectadas Caso 2 parte 02	52
C.3. Dimensiones relevantes detectadas Caso 2 parte 03	52
C.4. Dimensiones relevantes detectadas Caso 2 parte 04	55
C.5. Dimensiones relevantes detectadas Caso 2 parte 05	55
C.6. Dimensiones relevantes detectadas Caso 2 parte 06, 07, 08, 09, 10	55

Índice de Ilustraciones

2.1.	Ejemplificación de la maldición de la multi-dimensionalidad. La suma o llegada de nuevas dimensiones separa los puntos disminuyendo su densidad.	9
2.2.	Dendograma	11
2.3.	Conjunto de datos con cuatro clusters, donde una de las dimensiones no es relevante para cada cluster.	15
2.4.	Conjunto de datos graficados en pares de dos dimensiones. En las primeras dos imágenes se ven los cluster bien separados mientras los otros se encuentran mezclados.	15
2.5.	Ejemplo ilustrativo de como puede migrar una base de datos a medida que cambia el valor dentro del objeto.	17
2.6.	La primera imagen corresponde a la una división completa de un cluster en dos nuevos. La segunda imagen corresponde a una división parcial del cluster. Dimensión A corresponde al tiempo.	17
2.7.	La primera imagen corresponde a la una fusión de dos clusters en uno nuevo. La segunda imagen corresponde a una fusión de un cluster a otro existente aumentando el tamaño del existente. Dimensión A corresponde al tiempo.	18
2.8.	La primera imagen corresponde a una esparsión de un cluster. La segunda imagen corresponde a una contracción. La tercera es un ligero cambio de forma a uno más circular(uniforme).	18
4.1.	Tiempos promedios de ejecución de 10 muestras del Caso 01	38
4.2.	Tiempos promedios de ejecución de 10 muestras del Caso 02	39
4.3.	Tiempos promedios de ejecución de 10 muestras del Caso 03	39
4.4.	Resultados métrica de 20 muestras del Caso 02	41
4.5.	Resultados métrica de 20 muestras del Caso 03	41
C.1.	Tiempo de ejecución de llegada de 10 lotes Caso 02 Estático	53
C.2.	Tiempo de ejecución de llegada de 10 lotes Caso 02 Dinámico	53
C.3.	Tiempo de ejecución de llegada de 10 lotes Caso 01 Estático	54
C.4.	Tiempo de ejecución de llegada de 10 lotes Caso 01 Estático	54