# Práctica 6 ACSIC

## Marc Llobera Villalonga Grupo 202

**El objetivo de esta parte es la comprensión del concepto de caracterización de la carga. Para ello, se hará uso de la herramienta Weka.**

**De la monitorización de un sistema de almacenamiento, se ha obtenido se proporciona un fichero de datos llamado “data.txt”. En el fichero se almacenan tres columnas con la siguiente información:**

**- El tamaño del fichero accedido (en MB). Los valores que correspondan con “-1” quieren decir que el acceso al fichero ha fallado.**

**- La hora a la que se hizo el acceso. El valor 22 representan las 22h, el valor 01 representan las 1h (a.m.), etc.**

**- El ancho de banda consumido (en MS/s). Los valores de esta columna están entre 453 y 1355, por lo tanto, los valores de esta columna deberán ser tratados. Es decir, el valor crudo de “1258.84,”, corresponde con “1258,84”.**

**Con los datos proporcionados se pide caracterizar la carga haciendo uso del algoritmo de K- means y responder a las siguientes preguntas:**

**- Aplicando el algoritmo con 100 iteraciones y agrupando los datos en 3 clases, ¿qué resultados se obtienen? Muéstralo gráficamente.**

En los datos proporcionados y procesados tenemos tres características que son la hora, el tamaño del fichero, y el ancho de banda, de estas tres características la hora es la única que no nos proporciona ninguna información relevante por lo que a partir de ahora no trabajaremos con ella. Así caracterizando la carga como se nos pide y solamente haciendo caso de el tamaño (SIZE) y del ancho de banda (MBS) nos sale el siguiente gráfico.

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

Donde los colores muestran la diferencia del ancho de banda de cada muestra y dividiendo los ficheros en 3 clases que podríamos nombrar ficheros de velocidad lenta, media y rápida. A su vez podemos ver también en el eje del tamaño que hay 3 clases de ficheros bien diferenciados como son ficheros de tamaño pequeño, de gran tamaño, y de tamaño grande.

|  |  |  |  |
| --- | --- | --- | --- |
| **Representantes** | **ficheros rápidos** | **ficheros lentos** | **ficheros intermedios** |
| **SIZE** | 4805.7748 | 4754.1331 | 4624.8436 |
| **MBS** | 1204.1454 | 602.7856 | 902.8461 |

**- Con el mismo número de iteraciones y agrupando los datos en 5 clases, ¿qué resultados se obtienen? ¿Cómo difieren de los anteriormente obtenidos?**

**Gráfico, Gráfico de dispersión

Descripción generada automáticamente**

La única diferencia que nos da es la clasificación en clases obviamente ya que ahora tenemos 5 clases diferentes para la diferencia de ancho de banda de cada fichero y clasificando estos según 5 tipos de velocidades. Aún podemos ver también la diferencia clara de grupos de ficheros según el tamaño de estos.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Representantes** | **ficheros**  **rápidos** | **ficheros**  **intermedios-rápidos** | **ficheros**  **intermedios** | **ficheros**  **intermedios-lentos** | **ficheros**  **lentos** |
| **SIZE** | 4648.5732 | 4792.8688 | 4755.7865 | 4604.4884 | 4839.2968 |
| **MBS** | 1265.0141 | 1084.9059 | 904.2082 | 724.1673 | 543.2424 |

**- ¿Hay alguna característica especial en la carga proporcionada? Explícala con detalle.**

Podemos ver desde un principio que para el SIZE tenemos una enorme cantidad de ficheros que pesan muy poco, cerca de 0 incluso muchos directamente tienen tamaño 0 y después se ve como hay un gran salto hasta otro grupo de ficheros cuyo tamaño es intermedio en comparación al tamaño máximo de fichero que podemos encontrar en los datos por lo que sin necesidad de hacer *clustering* podemos hacer una clasificación clara de ficheros pequeños, medianos, y grandes en cuanto a tamaño, por lo que podemos separarlos en 3 ficheros de datos y realizar *cluster* sobre cada uno por separado y conseguir una mejor clasificación según las velocidades (MBS) sabiendo que unos ficheros son de tamaño menor (0 - 963961), otros de tamaño intermedio (963962 - 1927923), y finalmente un pequeño grupo de ficheros de gran tamaño (1927924 - 2891887). Así, nos quedarían estos gráficos para 3 clusters:

**Tamaño pequeño (0 - 963961)**

Gráfico

Descripción generada automáticamente

|  |  |  |  |
| --- | --- | --- | --- |
| **Representantes**  **ficheros tamaño pequeño** | **ficheros**  **rápidos** | **Ficheros**  **intermedios** | **ficheros**  **lentos** |
| **SIZE** | 3650.2762 | 3570.3511 | 3609.7 |
| **MBS** | 1204.2039 | 902.9535 | 602.8424 |

**Tamaño medio (963962 – 1927923)**

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

|  |  |  |  |
| --- | --- | --- | --- |
| **Representantes**  **ficheros tamaño medio** | **ficheros**  **rápidos** | **Ficheros**  **intermedios** | **ficheros**  **lentos** |
| **SIZE** | 1519992.9568 | 1506484.6762 | 1497968.6308 |
| **MBS** | 1121.5611 | 912.3837 | 658.894 |

**Gran tamaño (1927924 – 2891887)**

Imagen que contiene Gráfico

Descripción generada automáticamente

|  |  |  |  |
| --- | --- | --- | --- |
| **Representantes**  **ficheros tamaño grande** | **ficheros**  **rápidos** | **Ficheros**  **intermedios** | **ficheros**  **lentos** |
| **SIZE** | 2085105.4375 | 2082105.3913 | 2093981.625 |
| **MBS** | 1127.383 | 991.8393 | 658.8919 |

Para 5 clusters:

**Tamaño pequeño (0 - 963961)**

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Representantes**  **fichero tamaño pequeño** | **ficheros**  **rápidos** | **ficheros**  **intermedios-rápidos** | **ficheros**  **intermedios** | **ficheros**  **intermedios-lentos** | **ficheros**  **lentos** |
| **SIZE** | 3655.9247 | 3631.3665 | 3563.9358 | 3565.4655 | 3634.1404 |
| **MBS** | 1265.0126 | 1084.8939 | 904.2081 | 724.1745 | 543.2435 |

**Tamaño medio (963962 – 1927923)**

Gráfico

Descripción generada automáticamente

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Representantes**  **fichero tamaño medio** | **ficheros**  **rápidos** | **ficheros**  **intermedios-rápidos** | **ficheros**  **intermedios** | **ficheros**  **intermedios-lentos** | **ficheros**  **lentos** |
| **SIZE** | 1603990.8514 | 1496885.4894 | 1315432.6579 | 1515713.0833 | 1517701.6071 |
| **MBS** | 1124.5908 | 1115.4407 | 1021.6486 | 659.0917 | 655.9049 |

**Gran tamaño (1927924 – 2891887)**

Imagen que contiene Gráfico

Descripción generada automáticamente

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Representantes**  **fichero tamaño grande** | **ficheros**  **rápidos** | **ficheros**  **intermedios-rápidos** | **ficheros**  **intermedios** | **ficheros**  **intermedios-lentos** | **ficheros**  **lentos** |
| **SIZE** | 2041996 | 2337459.6364 | 2045823.9167 | 2055392.0909 | 2034378.5 |
| **MBS** | 1182.2457 | 1088.6163 | 826.5108 | 677.3882 | 675.3232 |