

IMPUTACIÓN DE INFORMACIÓN DE CONTROL FALTANTE EN LA GESTIÓN DE RECURSOS Y PROCESOS

Diego David Duré Attis¹, Jorge Tomás Fornerón Martínez¹, Federico Agostini²
y David Luis La Red Martínez²

¹Facultad de Ciencias Aplicadas, Universidad Nacional de Pilar, Pilar, Paraguay

²Facultad de Cs. Ex. y Naturales y Agrimensura, Universidad Nacional del Nordeste, Corrientes, Argentina

RESUMEN

Los sistemas distribuidos se componen de múltiples nodos que intercambian información entre sí, para mantener la red conectada, cada nodo es único e independiente y tiene una capacidad de procesamiento diferente a los demás. En esta propuesta se considera que hay un nodo central que se encarga de recibir y mantener actualizada la información de control de todos ellos, y es el encargado de la asignación de recursos en la modalidad de exclusión mutua, asegurando la disponibilidad de los mismos y respetando las prioridades de los procesos. En un determinado ciclo de recolección de información de gestión, necesaria para ejecutar y asegurar lo mencionado anteriormente, el nodo central puede recibir de alguno de los nodos información incompleta, por ejemplo, de los criterios de evaluación de carga nodal y de las prioridades o preferencias de los procesos. Estos datos faltantes, constituyen un obstáculo importante en la gestión de recursos. Las técnicas de imputación de datos permiten estimarlos utilizando diferentes algoritmos, mediante los cuales, se puede imputar una característica importante para una instancia en particular. Esta investigación propone el uso de imputación de valores faltantes sobre la información de control en el contexto de los Sistemas Distribuidos. Se incorpora una capa de imputación/asignación a un modelo de decisión, que permite completar los valores faltantes con valores estimados, necesarios para establecer un correcto orden de asignación de recursos a los procesos. Se utiliza el método de imputación K-Means, por ser considerado uno de los más fiables y de menor consumo de recursos computacionales.

PALABRAS CLAVE

Sistemas Operativos, Operadores de Agregación, Modelo de Decisión, Imputación de Datos, K-Means

1. INTRODUCCION

Los sistemas distribuidos están compuestos por un grupo de ordenadores cuyos componentes (hardware y software) están conectados en red, y les permite comunicarse y coordinar acciones para lograr un objetivo. En los sistemas informáticos distribuidos donde existen múltiples procesos que cooperan para el logro de una determinada función, es indispensable disponer de modelos de decisión que permitan a los grupos de procesos acceder a los recursos compartidos que sólo se pueden acceder en la modalidad de exclusión mutua. Estos procesos se encuentran distribuidos en múltiples nodos, en los cuales se define una interfaz entre las aplicaciones y el sistema operativo, que a través de un Runtime (software en tiempo de ejecución complementario al sistema operativo presente en cada uno de los nodos) incluido en esa interfaz, gestiona los procesos y recursos compartidos. Los Runtime interactúan entre sí para intercambiar información y uno de ellos actúa como coordinador global para las asignaciones.

En cierto momento alguno de los nodos del sistema distribuido puede ser considerado inaccesible, no pudiendo brindar información al Runtime central (coordinador), que a su vez, va postergando innecesariamente solicitudes de recursos por parte de los procesos a siguientes rondas de asignación. Esto ocurre específicamente por la falta de información de control sobre los requerimientos de los procesos que pertenecen a ese nodo.

Las principales fuentes de interés son las siguientes: a) Los métodos considerados tradicionales para la asignación de recursos en sistemas distribuidos respetando la exclusión mutua en el acceso a los recursos compartidos, según se expresa en Ricart & Agrawala (1981), Lodha & Kshemkalyani (2000); b) Los métodos considerados innovadores expresados en La Red Martínez (2017), La Red Martínez, Acosta & Agostini (2018),

Agostini & La Red Martínez (2019), Agostini, La Red Martínez & Acosta (2018), Agostini, La Red Martínez & Fornerón Martínez (2019) y Fornerón, Agostini & La Red Martínez (2020); c) Los principales métodos de imputación de datos descriptos en Jönsson & Wohlin (2004), Mehala & Vivekanandan (2008), Husson, Josse, Narasimhan & Robin (2019) y Choudhury & Pal (2019). Otros métodos se evalúan en Primorac, La Red Martínez & Giovannini (2020), Rahman & Davis (2013) y en Patil, Joshil & Toshniwal (2010); d) Los modelos de decisión utilizando operadores de agregación adecuados mostrados en Peláez, Doña & La Red Martínez (2004), Peláez, Doña & Gómez-Ruiz (2007), Dong, Zhang & Herrera-Viedma (2016); e) Los mecanismos de amputación que generan conjuntos de datos con valores faltantes a partir de conjuntos de datos completos expresados en Schouten, Lugtig & Vink (2018), Schouten & Vink (2018) y Primorac et al. (2020).

Por tal motivo se pretende, generar un método de imputación/asignación para estimar las variables de estados de los nodos considerados inaccesibles, que a través de un modelo de decisión, permita sustituir los datos faltantes por los valores de sustitución o valores imputados. Los valores faltantes en la información de estado de carga enviada por los nodos del sistema distribuido deben ser reemplazados por valores estimados aplicando un método de imputación, para que el modelo de decisión pueda establecer un correcto orden de asignación de recursos. Las técnicas de imputación de datos permiten estimar valores faltantes en el conjunto de datos utilizando diferentes algoritmos, mediante los cuales se puede imputar una característica importante para una instancia en particular.

La propuesta es presentar un método innovador para la gestión de recursos compartidos en sistemas distribuidos, basado en La Red Martínez (2017) y en Agostini et al. (2019), donde se desarrollan operadores de agregación para asignar recursos en sistemas distribuidos, pero que no consideran la posibilidad de información de control faltante. Este trabajo consiste en resolver esta problemática, agregando una capa de imputación/asignación de información de control, a un modelo de decisión existente para gestión de recursos y procesos en sistemas distribuidos, considerando como punto de inicio las premisas y las estructuras de datos mencionadas en esas publicaciones. Este trabajo se ha estructurado de la siguiente manera: en la sección 2 se presenta la propuesta de solución al problema mencionado, en la sección 3 se detalla el operador de agregación utilizado con las limitaciones que impone el número de páginas estipulado, en la sección 4 se indican las principales conclusiones y trabajos futuros, para finalizar con los agradecimientos y las referencias.

2. PROPUESTA DE SOLUCIÓN

En una macro imagen del sistema, se consideran transiciones de estados de acuerdo a la actualización iterativa de los requerimientos del sistema, en donde se encuentran los distintos nodos, cuyo estado puede ser: a) activo, cuando interactúa con los demás nodos aunque no envíe información ni participe del uso de recursos o la ejecución de procesos con recursos externos; b) activo compartido, cuando interactúa con los demás nodos, enviando y recibiendo información, creando procesos que utilizan recursos de otros nodos y compartiendo recursos propios con nodos externos; c) inactivo, cuando no se tiene información del estado real del nodo (puede estar sobrecargado y no puede realizar el intercambio de información, o bien, pudo haberse caído el enlace que lo conecta con los demás nodos). Por último, existe un nodo central, que a través de un Runtime gestiona la información de todos los nodos que participen en las asignaciones de recursos.

El nodo central, es el encargado de recopilar la información de todos los nodos, aplicar el proceso de agregación y obtener la lista de asignaciones de recursos a procesos. Un macro ciclo contempla la iteración de este proceso, en el cual el modelo de decisión es una etapa que se repite n veces. En este macro ciclo hay un paso previo de recopilación de información del estado del sistema, nodos activos y conectados, cuántos están entregando información, es decir un sondeo previo. Luego de la recepción se continúa con el proceso de evaluación y la emisión de los resultados. Se vuelve a aplicar el ciclo y el universo va cambiando en cada iteración. La falta de información en alguno de estos aspectos, será resuelta mediante los métodos comentados a continuación.

2.1 Método de Imputación K-Means

Es un algoritmo de clasificación no supervisada que agrupa a las observaciones de forma tal que todas las que se encuentren en el mismo grupo sean lo más semejantes entre sí, la semejanza se calcula con una métrica de similitud, calculando la media del grupo seleccionado por la semejanza (imputación con la media), eso evita el sesgo en la varianza y en la desviación estándar (estimación en base a los valores parecidos y no al total).

En K-Means la cantidad de valores parecidos surge de la clusterización realizada (si se tiene que imputar el valor de una variable de un registro de un grupo de 100 registros que componen el clúster, se tomarán los valores disponibles de esa variable a imputar dentro del grupo de los 100).

2.2 ¿Cómo Funciona el Algoritmo K-Means?

Según lo expuesto en Mehala & Vivekanandan (2008) K-Means define los centroides de k , uno para cada grupo. Estos centroides deben ser colocados de una manera a mejorar, porque una ubicación diferente causa un resultado diferente. Por lo tanto, la mejor opción es colocarlos lo más lejos posible uno del otro. El siguiente paso es tomar cada punto perteneciente a un grupo de datos y asociarlo al centroide más cercano. Cuando no hay ningún punto pendiente, el primer paso se completa y se hace un agrupamiento. En este punto hay que recalcular los nuevos centroides k como baricentros de los grupos resultantes del paso anterior. Después de estos nuevos centroides k , hay que hacer un nuevo agrupamiento entre los mismos puntos del grupo de datos y el nuevo centroide más cercano. Se ha generado un bucle. Como resultado de este bucle, puede que los centroides k cambien su ubicación paso a paso hasta que no se hagan más cambios. Calcula la media de todos los puntos del grupo. Consideremos que k es el número de grupo. Consideremos que el número k de grupos es el valor x_{ij} de la clase k -ésima, C_k falta en el grupo k entonces será reemplazado por:

$$x_{ij} = \sum_{i: x_{ij} \in C_k} \frac{x_{ij}}{n_k}$$

donde, n_k representa el número de valores no faltantes en la j -ésima parte de la clase k -ésima Patil et al. (2010).

La limitación de este método es que sustituirá dentro de cada agrupamiento todos los valores que faltan por el mismo valor, aunque la variabilidad definida está asociada a los datos reales que faltan. En consecuencia, se modifica la distribución estadística de los datos y ello repercutirá en la calidad de los mismos. Este es un enfoque simple y el más utilizado en la literatura. Una vez agrupados todos los registros, el K-Means calcula la media del grupo seleccionado por la semejanza, para obtener el valor imputado para una variable de un registro.

2.3 Aplicar el Método K-Means con el Lenguaje de Programación R

Sin pérdida de generalidad, para la imputación de datos en esta propuesta se utiliza el lenguaje de programación R, R Core Team (2020). El método de imputación seleccionado es el K-Means, éste imputa el valor faltante utilizando el lenguaje de programación mencionado, primeramente, calcula la cantidad óptima de clúster, y seguidamente utiliza la información de los clúster obtenidos para imputar los valores faltantes.

3. OPERADOR DE AGREGACIÓN UTILIZADO POR EL MODELO DE DECISION

En La Red Martínez (2017), las características de los operadores de agregación descritos permiten considerar que el método propuesto pertenece a la familia de operadores de agregación Neat-OWA. La propuesta desarrollada en el trabajo citado consiste en generar un modelo de decisión y su correspondiente operador de agregación para la gestión de grupos de procesos que comparten recursos, realizando modificaciones de los operadores de la familia OWA mencionados.

La propuesta mencionada anteriormente no considera la información de control faltante, por lo que en este trabajo se propone incorporar una capa de imputación/asignación, teniendo en cuenta los siguientes pasos: **a)** Identificar el propósito del problema; **b)** Identificar las alternativas; **c)** Listar los criterios que van a ser tenidos en cuenta a elegir la mejor alternativa.

3.1 Identificar el Problema

Este paso consiste en identificar qué es lo que se desea alcanzar. Para este trabajo lo que se intenta resolver son las solicitudes de recursos por parte de los procesos que se encuentran en nodos distribuidos. A partir de un nodo central que aloja al Runtime, se recibe la información de control respecto del estado de los nodos y de los

requerimientos de acceso de procesos a recursos, proporcionada por todos los nodos que están compartiendo información, y a través de un operador de agregación determina y soluciona la siguiente premisa: “que los procesos accedan a recursos compartidos en la modalidad de exclusión mutua, incorporando mecanismos de imputación de datos para los casos en que la información sobre las variables que indican el estado de carga de alguno o algunos de los nodos sea incompleta”.

En diversas situaciones, el nodo central puede recibir información incompleta sobre alguno de los nodos; por ejemplo, alguno de los criterios de evaluación de carga (ej.: CPU, memoria virtual, prioridad de proceso, etc.) puede no estar disponible. Como solución a este problema se pretende mejorar un modelo de decisión y su correspondiente operador de agregación estudiando la aplicación de imputación de datos cuando la información de las variables que indican el estado de carga de alguno o algunos de los nodos sea incompleta, para que el modelo de decisión pueda establecer un orden de asignación de recursos. En caso de que la información suministrada por los nodos esté completa, el modelo de decisión se ejecuta sin necesidad de la imputación de datos, según la propuesta de La Red Martínez (2017) y Agostini et al. (2019).

La imputación de datos será la herramienta a utilizar en el modelo de decisión propuesto para la asignación de procesos a recursos en sistemas distribuidos. En base a revisiones bibliográficas se opta por utilizar para este escenario el método K-Means, por ser uno de los métodos más utilizados, en razón de sus resultados, en investigaciones donde los resultados obtenidos son comparados con otros métodos.

3.2 Identificar las Alternativas

Son las opciones que podrían ayudar a la consecución del propósito que se trazó en el primer paso. El Runtime alojado en cada nodo gestiona los procesos y recursos compartidos, para que luego defina el escenario correspondiente. En un determinado ciclo de recolección de información proporcionada por todos los nodos, el nodo central puede recibir información incompleta de uno o algunos de los criterios utilizados para la carga computacional de cada nodo.

Además, para cada nodo, se consideran distintos criterios para calcular la prioridad o preferencia que tiene cada uno, en las asignaciones de recursos a procesos. Al igual que el caso anterior, algunos de los valores de estos criterios podrían no estar disponibles en un ciclo de recolección.

Para obtener un indicador de las prestaciones de cada nodo, se tendrá en cuenta sus características, por ejemplo, la velocidad de procesamiento, capacidad de memoria, velocidad de transmisión de datos, velocidad de entrada de salida, entre otros. Estas características permitirán identificar nodos de altas, media y bajas prestaciones. El objetivo principal será favorecer a los nodos de mejores prestaciones por sobre los demás, por considerar que esos nodos estarán en mejores condiciones para resolver las diferentes asignaciones con diferentes cargas de trabajo. Cuando un nodo arranca, informará a través de su Runtime, el detalle de sus características al Runtime del nodo central, permitiendo que este último pueda clasificarlos según sus prestaciones.

3.3 Listar los Criterios que van a ser Tenidos en Cuenta para Elegir la Mejor Alternativa

3.3.1 Cálculo de la Carga Computacional Actual de los Nodos

Para obtener un indicador de la carga computacional actual de cada nodo se adoptarán los mismos e criterios en los n nodos. Los valores que se asumirán para los indicadores de carga computacional de los n nodos y el cálculo de carga promedio para cada nodo serán los obtenidos por el nodo central. Puede presentarse la situación de que en cierto ciclo de recolección el nodo central reciba información incompleta de uno o algunos de los criterios (% de uso de CPU, % de uso de Memoria y/o el % de uso de Operación de Entrada/Salida) a utilizar para el cálculo de la carga computacional, entonces se aplica la imputación de datos para dichos criterios, para imputar se tendrá en cuenta ciertas pautas, sabiendo que son las variables de estados consideradas para los nodos.

En función a la información de control histórica (disponible en el nodo central y que corresponde a la información de carga computacional y preferencias de los procesos en ciclos anteriores) disponible para los métodos de imputación/asignación, se describirán distintas pautas para las diferentes categorías.

Primera pauta: para el criterio ij del nodo k , **si se cuenta** con información histórica, se puede realizar la imputación con K-Means. Para ello, se aplica la imputación de datos teniendo en cuenta un histórico de valores

correspondientes a cierta cantidad de ciclos de recolección de información del nodo k , se obtiene entonces, el valor faltante del criterio ij .

Segunda pauta: para el criterio ij del nodo k , **si no se cuenta** con información histórica, se puede realizar la asignación de un valor por defecto, en función de la carga computacional informada y las características del nodo:

- 1) Se calcula el promedio de la carga computacional existente basadas en la información disponible de los valores de criterios de carga computacional informados por el nodo central para el criterio ij del nodo k . El promedio carga computacional existente es igual a la suma de los valores disponibles de criterios de carga computacional dividido la cantidad de valores disponibles.
- 2) Se determina la prestación del nodo en base a las características conocidas. Los valores de las prestaciones mencionadas anteriormente, permiten determinar las características del nodo k , por lo tanto, para el criterio ij se asume el valor correspondiente.
- 3) Se promedia los valores del primer y segundo paso. El promedio del primer y segundo paso es igual al valor del promedio de la carga computacional existente sumado al valor de las prestaciones del nodo, todo ello, dividido por la cantidad de criterios disponibles.

3.3.2 Cálculo de las Prioridades o Preferencias de los Procesos Teniendo en Cuenta el Estado del Nodo

Puede presentarse que en un determinado ciclo de recolección de información de control el nodo central reciba de alguno o algunos nodos información incompleta sobre los criterios de evaluación de carga, que son los valores que permiten calcular las prioridades o preferencias de los procesos, entonces se aplica el método de imputación teniendo en cuenta para las variables, tres categorías establecidas con sus respectivas pautas.

La primera categoría incluye variables medidas por el sistema operativo que representan la carga del nodo, por ejemplo, las variables de estados principales del nodo, que son las medidas que se obtienen del sistema operativo cada vez que le llega un requerimiento. La segunda considera los valores asignados por defecto del sistema operativo o asignado externamente, por ejemplo, la prioridad del proceso. La tercera considera los valores que indican cuánto le costaría al nodo hacer la asignación, el impacto esperado de atender el requerimiento (se tiene en cuenta la relación de proceso-recurso).

En función a la información de control histórica disponible para los métodos de imputación/asignación, se describirán distintas pautas para las diferentes categorías.

Categoría 1

Primera pauta: para el criterio ij de la relación proceso-recurso $p_{kl}r_{ij}$, **si se cuenta** con información histórica, se puede realizar la imputación con K-Means. Para ello, se aplica la imputación de datos teniendo en cuenta un histórico de valores correspondientes a cierta cantidad de ciclos de recolección de información de la relación proceso-recurso $p_{kl}r_{ij}$, se obtiene entonces, el valor faltante del criterio ij .

Segunda pauta: Para el criterio ij de la relación proceso-recurso $p_{kl}r_{ij}$, **no se cuenta** con información histórica específica de esa relación, **pero sí se tiene información histórica del proceso con relación a otros recursos**, se puede realizar la imputación con K-Means. Para ello, se aplica la imputación de datos teniendo en cuenta un histórico de valores correspondientes a cierta cantidad de ciclos de recolección de información de la relación proceso p_{kl} con otros recursos r_{ij} , se obtiene entonces, el valor faltante del criterio ij .

Tercera pauta: Para el criterio ij de la relación proceso-recurso $p_{kl}r_{ij}$, **si no se cuenta** con información histórica, **tampoco se tiene información histórica del proceso con otros recursos**, se puede realizar la asignación de un valor por defecto, en función de la carga computacional informada y las características del nodo:

- 1) Se calcula el promedio de los valores de los criterios ij (correspondientes a la categoría 1), disponibles para la relación $p_{kl}r_{ij}$. El promedio de los criterios existentes es igual a la suma de los valores disponibles de criterios ij , dividido por la cantidad de criterios disponibles.
- 2) Se determina la prestación del nodo en base a las características conocidas. Los valores de las prestaciones mencionadas anteriormente, permiten determinar las características del nodo k , por lo tanto, para el criterio ij se asume el valor correspondiente.
- 3) Se promedia los valores del primer y segundo paso, esto es igual al valor del promedio de los criterios disponibles sumado al valor de las prestaciones del nodo, todo ello, dividido por la cantidad de criterios disponibles.

Categoría 2

Primera pauta: para el criterio ij de la relación proceso-recurso $p_{kl}r_{ij}$, **si se cuenta** con información histórica, se puede realizar la imputación con K-Means. Para ello, se aplica la imputación de datos teniendo en cuenta un

histórico de valores correspondientes a cierta cantidad de ciclos de recolección de información de la relación proceso-recurso $p_{kl}r_{ij}$, se obtiene entonces, el valor faltante del criterio ij .

Segunda pauta: Para el criterio ij de la relación proceso-recurso $p_{kl}r_{ij}$, **no se cuenta** con información histórica específica de esa relación, **pero sí se tiene información histórica del proceso con relación a otros recursos**, se puede realizar la imputación con K-Means. Para ello, se aplica la imputación de datos teniendo en cuenta un histórico de valores correspondientes a cierta cantidad de ciclos de recolección de información de la relación proceso p_{kl} con otros recursos r_{ij} , se obtiene entonces, el valor faltante del criterio ij .

Tercera pauta: Para el criterio ij de la relación proceso-recurso $p_{kl}r_{ij}$, **si no se cuenta** con información histórica, **tampoco se tiene información histórica del proceso con otros recursos**, y al **no tener información sobre otros criterios** (referente a la categoría 2), se asumirá un valor por defecto en función a las características del nodo. Este paso consiste en determinar las prestaciones del nodo k en base a las características conocidas.

Categoría 3

Primera pauta: Para el criterio ij de la relación proceso-recurso $p_{kl}r_{ij}$, **si se cuenta** con información histórica, se puede realizar la imputación con K-Means. Para ello, se aplica la imputación de datos teniendo en cuenta un histórico de valores correspondientes a cierta cantidad de ciclos de recolección de información de la relación proceso-recurso $p_{kl}r_{ij}$, se obtiene entonces, el valor faltante del criterio ij .

Segunda pauta: Para el criterio ij de la relación proceso-recurso $p_{kl}r_{ij}$, **si no se cuenta** con información histórica, **tampoco se tiene información histórica del proceso con otros recursos**, se puede realizar la asignación de un valor por defecto, en función de la carga computacional informada y las características del nodo:

- 1) Se calcula el promedio de los criterios ij (correspondientes a la categoría 3), disponibles para la relación $p_{kl}r_{ij}$. El promedio de los criterios existentes es igual a la suma de los valores disponibles de criterios ij , dividido por la cantidad de criterios disponibles.
- 2) Se determina la prestación del nodo en base a las características conocidas. Los valores de las prestaciones mencionadas anteriormente, permiten determinar las características del nodo k , por lo tanto, para el criterio ij se asume el valor correspondiente.
- 3) Se promedia los valores del primer y segundo paso, esto es igual al valor del promedio de los criterios disponibles sumado al valor de las prestaciones del nodo, todo ello, dividido por la cantidad de criterios disponibles.

Luego de aplicar la imputación de datos faltantes necesarios para tener toda la información que utiliza el modelo de decisión para resolver los distintos escenarios desarrollados en La Red Martínez (2017) se prosigue con los cálculos previstos en el mismo.

3.3.3 Caso de Estudio

Los datos de entrada, recopilados por el nodo central, son los mismos que en La Red Martínez (2017) . En la Tabla 1, a partir de la evaluación de la información de control obtenida, se realiza una clasificación de la misma, considerando o no, las distintas categorías y pautas, definiendo los escenarios teniendo en cuenta la categoría de carga computacional y las prioridades o preferencias nodales, respectivamente.

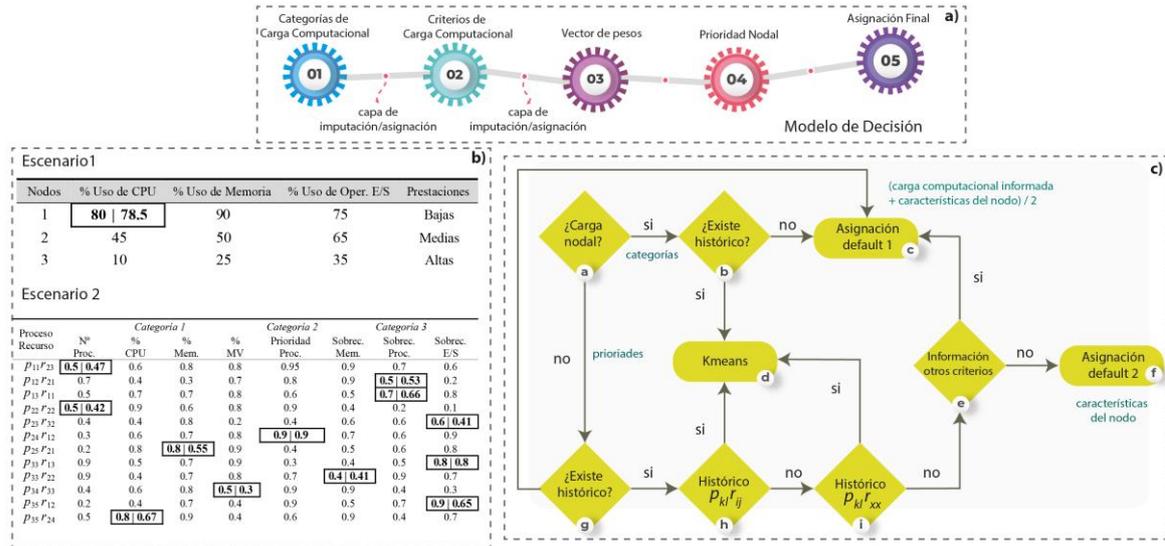
Tabla 1. Clasificación de datos faltantes en base a distintos escenarios

Nodo / Proceso-Recurso	Escenario	Categoría	Pauta 1	Pauta 2	Pauta 3	Pasos
<i>Nodo 1</i>	1		x			<i>a, b y d</i>
<i>p11r23, p35r24</i>	2	1	x			<i>a, g, h y d</i>
<i>p22r22, p34r33</i>	2	1		x		<i>a, g, h, i y d</i>
<i>p25r21</i>	2	1			x	<i>a, g, h, i, e y c</i>
<i>p24r12</i>	2	2	x			<i>a, g, h, d</i>
<i>p12r21, p13r11, p23r32, p33r13, p33r22</i>	2	3	x			<i>a, g, h, d</i>
<i>p35r12</i>	2	3		x		<i>a, g, h, i, e y c</i>

El color gris oscuro indica que no existe la categoría o pauta. El color gris claro indica que existe la pauta, pero no corresponde y el color blanco representa la pauta correcta. La columna de pasos, se corresponde con los pasos del diagrama de flujo de la Figura 1c.

En la Figura 1, se representa un ejemplo de aplicación del método desarrollado. En la Figura 1a, cabe destacar que los pasos del 1 al 5 corresponden al modelo de decisión tomado como referencia La Red Martínez (2017), al cual se le agrega una capa de resolución de datos faltantes (que se utilizará cuando sea necesario). Para ello, se inserta esta capa de “imputación/asignación” dentro del proceso general entre los pasos 1-2 y 2-3. Los escenarios mencionados en la Figura 1b, se pueden resolver mediante el diagrama de flujos de la Figura

1c. En las celdas remarcadas de la Figura 1b, el valor de la izquierda corresponde a los valores obtenidos en La Red Martínez (2017), y el de la derecha es el resultado de la imputación con K-Means, utilizando información histórica de mil ciclos.



3.3.4 Evaluación

Observando la poca diferencia que existe entre los valores originales y los imputados se determina que el método aplicado fue muy efectivo. A diferencia de otros métodos, ésta propuesta considera la información estimada (resultado de las imputaciones/asignaciones en los diferentes ciclos) como parte de la información histórica para aplicar una nueva imputación/asignación. Una de las limitaciones más importantes de esta propuesta es la inexistencia de información de control de alguno de los nodos, es decir, que falten la totalidad de datos correspondiente a un registro. Para resolver esto, se utilizará un nuevo mecanismo que utiliza medias ponderadas, en los cuales los valores de los registros más recientes tendrán mayor peso que los demás. Se aplicará el método K-Means siempre y cuando existan más de 10 registros históricos y falten algunos de los criterios mencionados, caso contrario, se aplica un método de medias ponderadas.

4. CONCLUSIONES Y TRABAJOS FUTUROS

El modelo de decisión propuesto considera la imputación de datos cuando la información sobre las variables que indican el estado de carga de alguno o algunos de los nodos llega incompleta, reemplazando esos valores faltantes por valores estimados, para que finalmente el modelo de decisión establezca un correcto orden de asignación de recursos a los procesos, respetando la exclusión mutua. Es importante mencionar que la solución presentada es compatible con la gestión de transacciones globales habitualmente implementada en los sistemas de gestión de base de datos.

Se prevé mejorar la evaluación de la propuesta considerando características como tiempo de ejecución, consumo de memoria y de CPU. Del mismo modo, dicha evaluación, será aplicada y comparada con otras aproximaciones que solventen el mismo problema.

Se tiene previsto considerar criterios de optimización del consumo de energía por parte de los componentes de los sistemas distribuidos de procesamiento en función de los objetivos de Desarrollo Sostenible de las Naciones Unidas y evaluar la migración controlada de procesos considerando la transferencia de procesos entre nodos de forma automática para hacer mejor uso de los recursos, característica que también está ligada a la distribución y el equilibrio de la carga de trabajo entre los nodos, contribuyendo a aumentar la tolerancia a fallos y así mejorar el rendimiento global del sistema distribuido, aportando a su autorregulación.

AGRADECIMIENTOS

Este trabajo ha sido apoyado por el Proyecto: "Modelos de decisión para la gestión de recursos y procesos en sistemas distribuidos considerando migración de procesos, imputación de datos y lógica difusa en los nuevos operadores de agregación", código 20F005 de la Universidad Nacional del Nordeste (Argentina).

REFERENCIAS

- Agostini, F., & La Red Martínez, D. L. (2019). Allocation of Shared Resources. 14th Iberian Conference on Information Systems and Technologies - CISTI 2019, 1–6.
- Agostini, F., La Red Martínez, D. L., & Acosta, J. C. (2018). Modeling of the consensus in the allocation of resources in distributed systems. *International Journal of Advanced Computer Science and Applications*, 9(12). doi: 10.14569/IJACSA.2018.091204.
- Agostini, Federico, La Red Martínez, D. L., & Fornerón Martínez, J. T. (2019). Nuevo Operador De Agregación Para Grupos De Procesos. *Conferências IADIS Ibero-Americanas WWW/Internet e Computação Aplicada 2019*, 223–230. doi:10.33965/ciaca2019_2019141028.
- Choudhury, S. J., & Pal, N. R. (2019). Imputation of missing data with neural networks for classification. *Knowledge-Based Systems*, 182. doi: 10.1016/j.knosys.2019.07.009.
- Dong, Y., Zhang, H., & Herrera-Viedma, E. (2016). Consensus reaching model in the complex and dynamic MAGDM problem. *Knowledge-Based Systems*, 106. doi: 10.1016/j.knosys.2016.05.046.
- Fornerón, J. T., Agostini, F., & La Red Martínez, D. L. (2020). Modelo de Decisión para Gestión de Procesos y Recursos en Sistemas Distribuidos con Balanceo de Carga de Trabajo. *Revista Ibérica de Sistemas y Tecnologías de Información (RISTI) - Iberian Journal of Information Systems and Technologies*, 353–366.
- Husson, F., Josse, J., Narasimhan, B., & Robin, G. (2019). Imputation of Mixed Data With Multilevel Singular Value Decomposition. *Journal of Computational and Graphical Statistics*, 28(3). doi: 10.1080/10618600.2019.1585261.
- Jönsson, P., & Wohlin, C. (2004). An evaluation of k-nearest neighbour imputation using Ilkert data. *Proceedings - International Software Metrics Symposium*. doi: 10.1109/METRIC.2004.1357895.
- La Red Martínez, D.L., Acosta, J. C., & Agostini, F. (2018). Assignment of resources in distributed systems. *IMCIC 2018 - 9th International Multi-Conference on Complexity, Informatics and Cybernetics, Proceedings*, 2.
- La Red Martínez, David L. (2017). Aggregation Operator for Assignment of Resources in Distributed Systems. *International Journal of Advanced Computer Science and Applications*, 8(10). doi: 10.14569/ijacsa.2017.081053.
- Lodha, S., & Kshemkalyani, A. (2000). A fair distributed mutual exclusion algorithm. *IEEE Transactions on Parallel and Distributed Systems*, 11(6). doi: 10.1109/71.862205.
- Mehala, B., & Vivekanandan, K. (2008). An Analysis on K-Means Algorithm as an Imputation Method to Deal with Missing Values. *Asian Journal of Information Technology*, 7, 434–441.
- Patil, B. M., Joshi, R. C., & Toshniwal, D. (2010). Missing value imputation based on k-mean clustering with weighted distance. *Communications in Computer and Information Science*, 94 CCIS(PART 1), 600–609. doi: 10.1007/978-3-642-14834-7_56.
- Peláez, J., Doña, J. M., & Gómez-Ruiz, J. A. (2007). Analysis of OWA operators in decision making for modelling the majority concept. *Applied Mathematics and Computation*, 186(2). doi: 10.1016/j.amc.2006.07.161.
- Peláez, J., Dona, J., & Red, D. La. (2004). Majority opinion in group decision making using the qma-owa operator. 449–454.
- Primorac, C. R., La Red Martínez, D. L., & Giovannini, M. E. (2020). Metodología De Evaluación Del Desempeño De Métodos De Imputación Mediante Una Métrica Tradicional Complementada Con Un Nuevo Indicador. *European Scientific Journal ESJ*, 16(18). doi: 10.19044/esj.2020.v16n18p61.
- R Core Team. (2020). *The R Project for Statistical Computing*, Vienna, Austria. <https://www.R-Project.Org/>.
- Rahman, M. M., & Davis, D. N. (2013). Machine learning-based missing value imputation method for clinical datasets. *Lecture Notes in Electrical Engineering*, 229 LNEE, 245–257. doi: 10.1007/978-94-007-6190-2-19.
- Ricart, G., & Agrawala, A. K. (1981). An optimal algorithm for mutual exclusion in computer networks. *Communications of the ACM*, 24(1). doi: 10.1145/358527.358537.
- Schouten, R. M., Lugtig, P., & Vink, G. (2018). Generating missing values for simulation purposes: a multivariate amputation procedure. *Journal of Statistical Computation and Simulation*, 88(15), 2909–2930. doi: 10.1080/00949655.2018.1491577.
- Schouten, R. M., & Vink, G. (2018). The Dance of the Mechanisms: How Observed Information Influences the Validity of Missingness Assumptions. *Sociological Methods and Research*. doi: 10.1177/0049124118799376.