

## Método de fusión de datos de fuentes heterogéneas para mantener la consistencia de datos

Julio Muñoz<sup>1</sup>, Guillermo Molero–Castillo<sup>1,2</sup>, Edgard Benítez–Guerrero<sup>1</sup>

<sup>1</sup> Universidad Veracruzana, Xalapa, Veracruz, México

<sup>2</sup> CONACYT–Universidad Veracruzana, Xalapa, Veracruz, México

{juliomunoz, edbenitez}@uv.mx, ggmolero@conacyt.mx

**Resumen.** Los sistemas conscientes del contexto utilizan datos obtenidos de diversas fuentes para adaptarse y proveer servicios de interés al usuario de acuerdo a sus necesidades, localización o interacción con el ambiente. Sin embargo, el uso de fuentes heterogéneas crea un amplio volumen de datos que pueden diferir en formato, velocidad de transmisión y pueden ser afectados por el ruido ambiental. Esto genera cierta inconsistencia en los datos, la cual debe ser detectada a tiempo para evitar análisis erróneos. Para esto se hace uso de la fusión de datos, que es la acción de integrar fuentes diversas para ser analizadas de acuerdo a un contexto determinado. En este trabajo se presenta el diseño conceptual de un método de fusión de datos de fuentes heterogéneas, obtenidas de información contextual, con el propósito de mantener la consistencia de los datos en el proceso de fusión (extracción, pre–procesamiento, fusión y carga de datos).

**Palabras clave:** fusión de datos, fuentes heterogéneas, inconsistencia de datos, método de fusión, sistemas conscientes del contexto.

### Data Fusion Method of Heterogeneous Sources to Maintain Data Consistency

**Abstract.** Context–aware systems use data obtained from various sources to adapt and provide services of interest to the user according to their needs, location or interaction with the environment. However, the use of heterogeneous sources creates a large volume of data that may differ in format, transmission speed and may be affected by environmental noise. This generates some inconsistency in the data, which must be detected in time to avoid erroneous analysis. This is done using data fusion, which is the action of integrating diverse sources to be analyzed according to a determined context. This paper presents the conceptual design of a data fusion method of heterogeneous sources, obtained from contextual information, with the aim of maintaining the consistency of the data during the fusion process (extraction, pre–processing, fusion, and loading data).

**Keywords:** context-aware systems, data inconsistency, data fusion, fusion method, heterogeneous sources.

## 1. Introducción

En los últimos años, la tecnología ha evolucionado y se ha adaptado a las necesidades de los usuarios, así como al desarrollo de nuevas técnicas de comunicación y a la aparición de nuevos dispositivos con alto poder de cómputo y de tamaño cada vez más pequeños [1]. Este desarrollo tecnológico en la actualidad apertura nuevos temas de interés relacionados con los sistemas de cómputo y la forma en que perciben, capturan y analizan los datos del entorno que los rodea. A esto se le conoce como contexto, el cual es el entorno físico, emocional y social en el que se encuentra inmerso el usuario y que le dan significado, sentido y valor a las acciones o actividades que se realizan a su alrededor [1, 2].

Mediante el contexto puede caracterizarse la situación de una entidad, como persona, lugar u objeto, considerada relevante para la interacción entre un usuario y un sistema [3]. Para analizar esta interacción es importante tener información del contexto, que responda algunos cuestionamientos, conocidos como las cinco Ws [4]: a) quién (who), que se enfoca en la identidad del usuario; b) qué (what), que hace referencia a lo que el usuario está haciendo; c) dónde (where), que es la localización del usuario; d) cuándo (when), asociado al tiempo; y e) por qué (why), que incluye elementos del estado emocional de la persona.

La acción de contextualizar la información implica poner en contexto una situación que es percibida, de manera aislada o conjunta, de todos aquellos elementos que la rodean y que influyen sobre esta acción. A esto se le conoce como cómputo consiente del contexto o sistemas consientes del contexto, que detectan la actividad del usuario y reaccionan a los cambios del mismo para proveerle servicios que le sean de utilidad en la realización de sus actividades cotidianas [1, 5].

Para hacer esta detección y reacción se necesita analizar diferentes fuentes de datos, ya sean provenientes de sensores físicos o fuentes lógicas [1]. Los sensores físicos como los acelerómetros miden el desplazamiento y posición, los fotodiodos miden la proximidad e intensidad de la luz, las termo–resistencias miden la temperatura, entre otros. Mientras que las fuentes lógicas proveen información no física, como: agenda del usuario, preferencias, configuraciones, entre otros. En general, estas fuentes de datos son heterogéneas debido a la variedad de los formatos y velocidades de captura. Estas características hacen que los datos obtenidos sean completamente diferentes [6].

Una actividad compleja, previo al análisis de la detección y reacción a los cambios en un determinado contexto, es la fusión de datos de fuentes heterogéneas. Esta fusión de datos es el proceso de detección, asociación, correlación, estimación y combinación de datos en varios niveles [7], que provienen de diferentes fuentes, como [6, 8]: sensores, bases de datos, bitácoras, observaciones, señales e incluso decisiones.

A la fecha, la fusión de datos ha sido empleada en diversas áreas, como [6, 7]: procesamiento de señales, teoría de la información, estimación, estadística, inferencia e inteligencia artificial; teniendo mayores avances en aplicaciones militares mediante el reconocimiento automático de objetivos, navegación de vehículos autónomos,

sensado remoto e identificación de amenazas. Otras aplicaciones no militares son el monitoreo de procesos industriales, robótica, aplicaciones médicas, entre otros.

Indudablemente, el interés por la fusión de datos es cada vez mayor debido a la creciente incorporación de sensores en los dispositivos y sistemas de cómputo, el objetivo es tener información útil como apoyo en el proceso de la toma de decisiones sobre un determinado suceso, objeto o acción [5, 6]. Recientemente se está empleando también la fusión de datos para integrar fuentes variadas para hacer detecciones y clasificaciones de actividades en hogares inteligentes [9], ambientes virtuales inmersivos [10], interfaces tangibles [11] y escritorios inteligentes [12], apoyo en la realización de actividades del usuario [4], fusión de datos de observaciones humanas [13], entre otros.

Sin embargo, debido a la interacción que hace el usuario con el sistema consiente del contexto se produce una amplia variedad de datos heterogéneos producto del uso de sensores multimodales, objetos y aplicaciones que tienen como propósito facilitar la eficiencia del trabajo del usuario, adaptándose a los cambios de contexto y caracterización de la situación [14]. Hacer la fusión de datos de estas fuentes heterogéneas constituye un importante reto, esto debido al amplio conjunto de sensores y otras fuentes de datos que se utilizan para llevar a cabo una interacción dinámica del usuario, los objetos y el entorno [1, 14].

En este artículo se presenta el diseño conceptual de un método de fusión de datos de fuentes heterogéneas, obtenidas de información contextual, con el propósito de mantener la consistencia de los datos en el proceso de fusión. Las fases que contempla el método son: extracción, pre-procesamiento, fusión y carga de datos. El documento está estructurado en tres partes principales: la primera son los antecedentes del trabajo de investigación, la segunda es el método del trabajo científico, y la tercera es la propuesta de solución del método de fusión como diseño conceptual.

## **2. Antecedentes**

Los primeros trabajos relacionados con los métodos de fusión de datos se remontan a 1786 con el método Condorcet, quien utilizó este enfoque para las votaciones en los modelos de la democracia [15]. Posteriormente, los métodos de fusión de datos se siguieron aplicando en diversas disciplinas, como: fiabilidad [16], reconocimiento de patrones [17], redes neuronales artificiales [18], proceso de toma de decisiones [8, 19], estimación estadística [20, 21], y predicción climática [22]. Además, actualmente una de las áreas que mayor impulso ha dado a la fusión de datos es la ciencia de datos, donde se emplean métodos para extraer, transformar y cargar fuentes de datos como parte del proceso de ingeniería de datos, previo a la analítica de éstos [3, 23].

En los sistemas consientes del contexto, los métodos de fusión de datos han demostrado ser especialmente útiles [7], debido a que proporcionan la posibilidad de utilizar los datos de múltiples fuentes. En este sentido, la fusión de los datos es un aspecto clave y crítico en sistemas con diversas fuentes (sensores, usuarios, entre otras). El objetivo es la acción de fusionar o combinar, de una forma eficiente, los datos de múltiples fuentes para superar las limitaciones de utilizar una sola fuente.

En el caso concreto del cómputo consiente del contexto, donde se obtienen datos de diversas fuentes que intentan describir las acciones que se generan alrededor del sistema o del usuario, existe una amplia red de sensores distribuidos de forma lógica, espacial o geográfica en un entorno y conectados por una red de transmisión. Los sensores pueden ser visuales (cámaras), auditivos (micrófonos), infrarrojos, sensores de humedad, temperatura, entre otros. Existen ventajas derivadas de utilizar sistemas de múltiples fuentes frente a sistemas tradicionales [24, 25]. Un ejemplo de fusión de datos en un sistema consiente del contexto se presenta en la Figura 1, donde los datos se obtienen de diversos sensores ( $S_1, S_2, \dots, S_n$ ), así como de fuentes variadas, como: bases de datos, almacenes de datos, bitácoras, agendas, entre otras.

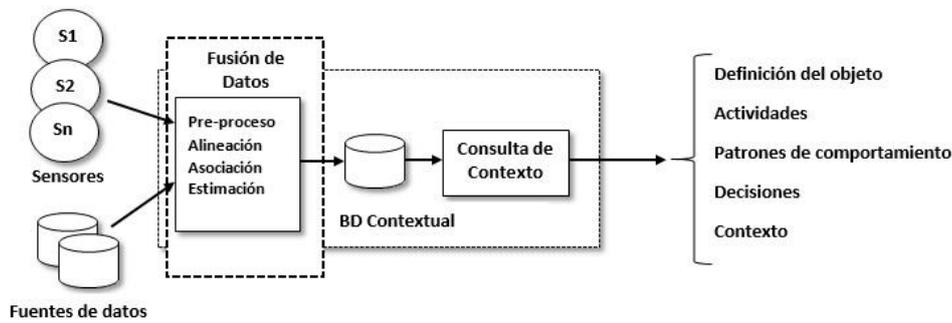


Fig. 1. Fusión de datos en un sistema consiente del contexto. Adaptada de [2].

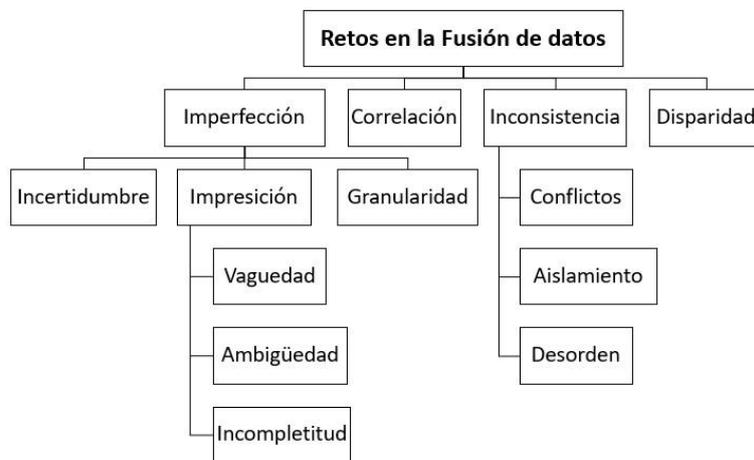


Fig. 2. Clasificación de los retos actuales en la fusión de datos. Adaptada de [31].

Los datos adquiridos se fusionan mediante un proceso de pre–procesamiento, alineación, asociación y estimación, y a través de consultas son comparados con una base de datos contextual, esto con el propósito de obtener una determinada inferencia del contexto. Esta inferencia proporciona información sobre la identidad del objeto, usuario, patrones de comportamiento o incluso permite al sistema tomar decisiones

adaptándose a la situación en un determinado momento, es decir, se integra la información de entrada (datos) para obtener datos refinados o características que describen las acciones del usuario y su comportamiento.

Así, para la implementación del proceso de fusión de datos existen diversos modelos, como Joint Directors Laboratories (JDL) [7], Thomopoulos [26], Integración de multi-sensores [27], Basado en conocimiento del comportamiento [28], Cascada [29] y Omnibus [30], y arquitecturas, como centralizada, descentralizada y distribuida [7], los cuales definen los niveles de complejidad, procesos y el momento cuando los datos deben ser fusionados. Sin embargo, debido a la naturaleza de los datos, a la amplia variedad de sensores y fuentes, y a la inexistencia de algoritmos ideales de fusión de datos, en [31] se propone una clasificación de retos actuales en la fusión de datos: a) imperfección, b) correlación, c) inconsistencia, y d) disparidad. Cada uno de estos retos son descritos en la sección siguiente.

## **2.1. Retos en la fusión de datos**

Los retos en la fusión de datos surgen debido al variado formato, tipo y velocidad de muestreo de los sensores, la diversidad de las fuentes y la imperfección de los datos. En la Figura 2 se presenta los retos identificados, los cuales fueron clasificados de acuerdo a su naturaleza [31]:

- Imperfección. Los datos suministrados por los sensores pueden ser afectados por cierto nivel de imprecisión, así como por la incertidumbre en las mediciones. Esta incertidumbre se presenta no solo por la imprecisión o el ruido de las mediciones de los sensores, sino también por las ambigüedades y la incapacidad del sistema de fusión de distinguirlas.
- Correlación. Este tipo de problema es común en configuraciones distribuidas, donde algunos de los sensores obtienen los mismos datos, de rutas diferentes o debido a rutas cíclicas del flujo de los datos.
- Inconsistencia. Constituye uno de los problemas más importantes en la fusión de datos debido a la incertidumbre inherente de las mediciones de los sensores, obteniéndose datos incorrectos. La causa de obtener estos datos incorrectos puede ser debido a fallas permanentes o fallas de corta duración en las mediciones.
- Disparidad. Los datos deben ser transformados localmente por cada sensor en un formato común, antes del proceso de fusión. Este problema de disparidad se presenta a menudo en los sensores debido a errores de calibración.

En la actualidad, cubrir todos los retos y limitaciones en la fusión de datos no es trivial, puesto que se produce una amplia variedad de datos heterogéneos debido al uso de sensores multimodales, objetos y aplicaciones que tienen como propósito facilitar la eficiencia del trabajo del usuario, adaptándose a los cambios de contexto y caracterización de una situación. Por tanto, la fusión de datos de múltiples fuentes es un aspecto clave y crítico en sistemas con variados sensores, usuarios y otros actores.

En este sentido, debido a la naturaleza de los datos existen trabajos que proponen soluciones parciales, pero no se tienen métodos que cubran los diferentes retos. Por lo que, para esta investigación se considera útil incluir para la fusión de datos métodos de fusión de fuentes heterogéneas, con el propósito de minimizar la inconsistencia en la fusión de datos contextuales. Así, se pretende obtener resultados con mejor calidad, a partir de múltiples sensores y fuentes variadas, realizando combinaciones de éstas de forma eficiente.

## **2.2. Inconsistencia de datos**

Los datos que provienen de sensores son afectados por cierto grado de imprecisión e incertidumbre en las mediciones. Esto genera inconsistencia en los datos. Además, las fuentes pueden tener diferentes velocidades de comunicación, esto provoca que los datos sean enviados de diversas maneras, sin mantener un orden al momento de ser recibidos, incluso pueden obtener valores atípicos debido a situaciones inesperadas, fallas en la medición y mediciones erróneas.

Algunos métodos utilizados para minimizar la inconsistencia de los datos son [4, 13, 31]: teoría de razonamiento evidencial Dempster Shafer, lógica difusa, estimación Bayesiana, filtro de Kalman, entre otros. Para la fusión de datos estos métodos toman en cuenta los factores ambientales y la imprecisión de los sensores que afectan las mediciones, así como la ambigüedad de los datos y la dificultad para distinguirlos [32]. Estas imprecisiones y ambigüedades generan inconsistencias que deben ser identificadas y tratadas para evitar errores durante el proceso de fusión, tales como [31]:

- **Conflicto en los datos.** Son datos atípicos recibidos por el sistema de fusión, los cuales deben ser descartados para evitar resultados erróneos. Estos pueden ser datos corruptos por factores ajenos al sistema de fusión, o provenir de fuentes con diversos formatos y dimensiones que necesitan ser integrados. Ante esto, se debe contar con mecanismos de estimación para prevenir esta clase de conflictos.
- **Datos en desorden.** Los datos para la fusión por lo general están organizados por la marca de tiempo que se les asigna cuando éstos fueron creados. Factores como la variación en los tiempos de propagación y fuentes heterogéneas hacen que los datos lleguen en una secuencia diferente a la esperada por el sistema de fusión. El principal problema es utilizar datos obsoletos para actualizar estados presentes, es decir, utilizar datos obtenidos con cierto retraso para la actualización de estados en tiempo real.
- **Aislamiento de los datos.** Esta clase de datos son mediciones erróneas por algún fallo en los sensores o el ruido de la señal provocado por el ambiente. Estos datos deben ser aislados para evitar estimaciones erróneas al ser fusionados con datos correctos. El objetivo es identificar o predecir estos datos para eliminarlos antes del proceso de fusión.

Por tanto, un aspecto fundamental en la fusión de datos es la capacidad de los métodos de fusión para mantener la consistencia de los datos, evitando conflictos,

aislamiento y desorden, provenientes de fuentes diversas, integrándolos de una manera adecuada, robusta y consistente [33] para obtener una visión unificada del evento, fenómeno o suceso monitoreado [34]. Sin embargo, en la fusión de datos de fuentes heterogéneas existen variados problemas que surgen durante el proceso de integración, tales como la asociación de los datos, la incertidumbre de los sensores provocada por el ruido inducido, esto debido a factores externos del ambiente, o incluso por imperfecciones en los mismos sensores, así como la administración de los datos, entre otros [33, 35].

Precisamente, ante los variados problemas en el proceso de fusión de datos, en la actualidad existen algunos estudios que se enfocan en la validación de sensores, antes al proceso de fusión; así como en el uso de conocimiento previo sobre los datos. Sin embargo, por lo general no siempre se tiene esta información debido a las condiciones reales en que opera el sistema de fusión [7, 31, 32], esto es, no es posible modelar en su totalidad todas las fallas inherentes a los sensores y fuentes variadas.

### 3. Método

Dado el propósito de mantener la consistencia de datos en el proceso de fusión de fuentes heterogéneas, para esta investigación se definieron cuatro etapas de trabajo; las cuales son de tipo exploratoria y aplicada que en su conjunto forman el método definido para esta investigación.



Fig. 3. Método de trabajo.

La primera es el inicio o preparación, la cual es una etapa fundamental que sustenta el análisis teórico de la fusión de datos y sus retos actuales. La segunda es el análisis y diseño conceptual de la propuesta de solución para la caracterización del método de fusión de datos de fuentes heterogéneas. La tercera etapa es el desarrollo del método de fusión con base en la elección de técnicas adecuadas aplicadas a un caso de estudio. La

cuarta es la evaluación del método de fusión de datos desde el punto de vista del desempeño y el cumplimiento del objetivo de mantener la consistencia de los datos en el proceso de fusión. En la Figura 3 se hace una proyección del método científico que se utiliza en esta investigación y las actividades contempladas en éste.

En virtud de lo anterior, en este artículo se presenta un avance de la investigación que cubre las primeras dos etapas del método de trabajo: i) inicio y ii) análisis y diseño. Ambas etapas fueron fundamentales como proceso analítico para diseñar esta propuesta conceptual del método de fusión de datos de fuentes heterogéneas, cuyo objetivo es mantener la consistencia de los datos en el proceso de fusión.

#### 4. Propuesta

Para los sistemas conscientes del contexto, el objetivo es incrementar la sensación de interacción del usuario con el sistema, ya sea mediante la manipulación física de objetos que afecten el comportamiento del sistema [36] o mediante la interacción y relación del usuario con el ambiente para mejorar el trabajo colaborativo con otros usuarios y dar soporte durante el proceso de la toma de decisiones [4]. Estas interacciones son detectadas mediante el uso de diversos sensores de video y proximidad, cámaras, acelerómetros, transceptores (receptores y transmisores), entre otros. El sistema reacciona a estas interacciones que representan información digital y adecúa su comportamiento como sea necesario, incluso creando interacciones físicas que respondan a las iniciadas por el usuario.

En este sentido, el uso de sensores diversos es importante para que estas interacciones puedan realizarse de manera adecuada, por lo que se debe tener un esquema de fusión de datos que permita explotar estos beneficios. La propia diversidad de los sensores y la naturaleza de las interacciones del usuario con el sistema pueden crear inconsistencia en los datos. Siendo estos datos necesarios para la caracterización del contexto.

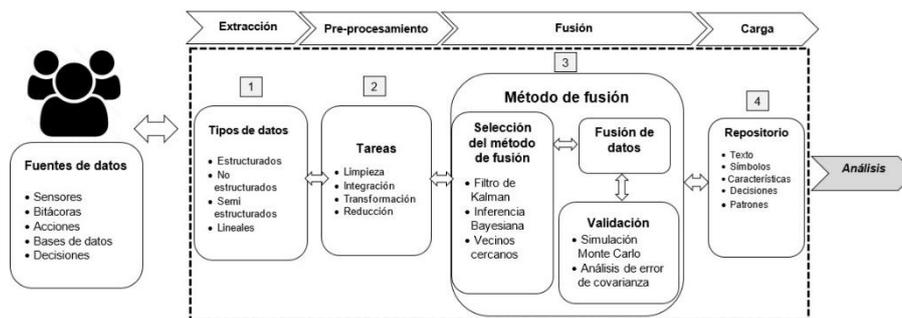


Fig. 4. Diseño conceptual del método de fusión de datos de fuentes heterogéneas.

La inconsistencia de datos es un factor crítico que debe ser detectado y evitado en el proceso de fusión. Esta fusión de datos inconsistentes con datos correctos puede producir una estimación errónea, que eventualmente puede conducir a un análisis

incorrecto [5]. Ante esto, un aspecto fundamental en la fusión de datos es la capacidad de los métodos de fusión para mantener la consistencia de los datos [31], evitando conflictos, aislamiento y desorden.

En este trabajo se propone el diseño conceptual de un método de fusión de datos de fuentes variadas (Figura 4), donde las interacciones del usuario con otros usuarios, con el ambiente o con el sistema generan datos heterogéneos obtenidos de sensores, bases de datos, agendas, configuraciones, preferencias, por mencionar algunas. Estas interacciones pueden ser acciones que realiza el usuario en un determinado espacio de trabajo, ambiente o entorno en el que se encuentra, e incluso pueden ser mediciones fisiológicas, como frecuencia respiratoria, cardíaca, temperatura, entre otros.

El diseño de este método de fusión está estructurado en cuatro etapas o módulos de procesamiento (extracción, pre-procesamiento, fusión y carga) que tiene como propósito principal mantener la consistencia de los datos a lo largo del proceso de fusión.

En la etapa de *extracción* los datos son obtenidos de diversas fuentes de acuerdo a su naturaleza, como sensores, bases de datos, bitácoras, acciones, mediciones fisiológicas, entre otras. Todas éstas con formatos diversos o incluso sin un formato establecido. Entre los tipos de datos considerados para procesar en este módulo están:

- Estructurados. Son datos que pueden estar almacenados en bases de datos, tablas, matrices, arreglos, bitácoras, entre otros.
- No estructurados. Son datos que no tienen un formato específico, como documentos de texto, datos crudos de sensores, redes sociales, entre otros.
- Semi-estructurados. Son datos que no están limitados a ciertos campos o tablas, pero tienen una forma de separar los elementos que los componen, como XML (eXtensible Markup Language), JSON (Java Script Object Notation), entre otros.
- Datos lineales. Estos datos están compuestos de texto plano, como archivo de texto delimitado (txt), valores separados por coma (csv), entre otros.

El *pre-procesamiento* constituye una etapa clave, previo a la fase de fusión de datos, puesto que los datos son preparados para ser fusionados y parte de la inconsistencia puede ser resuelta o minimizada en esta etapa. Para esto fueron definidas cuatro tareas generales:

- Limpieza. Consiste en hacer un proceso de limpieza o depuración para evitar datos incompletos, duplicados o conflictivos. Esto con el fin de obtener datos de mayor calidad, minimizando la inconsistencia de éstos.
- Integración. Consiste en agrupar los datos que vienen de diversas fuentes, aunque éstos tengan formatos diferentes. El objetivo es obtener fuentes de datos coherentes y consistentes, evitando problemas de representación y codificación.
- Transformación. Consiste en transformar valores numéricos a nominales o viceversa, ya sea mediante discretizaciones o derivaciones. Esta última permite crear nuevos atributos, incluso cambiando el formato de los datos. Además,

pueden hacerse normalizaciones para eliminar variaciones en los datos ajustándolos a una escala específica.

- Reducción. Consiste en reducir la alta dimensionalidad de los datos. En este caso el tamaño de éstos puede ser a consecuencia de datos redundantes. Para esto se obtienen muestras o representaciones de los datos, manteniendo su integridad, o usando mecanismo de reducción como análisis de componentes principales y análisis correlacional de datos.

Posterior al pre–procesamiento, en la fase de *fusión* se integran los datos con base en la selección de uno o más algoritmos. La selección del algoritmo depende de la naturaleza, el tipo y el nivel de abstracción de los datos, como: datos, atributos, símbolos, identidad, patrones o decisiones. Estos algoritmos deben ser capaces de distinguir la inconsistencia y combinar diferentes tipos de datos, a diferentes niveles, para obtener una descripción consistente [13, 32]. Por lo que, para este trabajo los algoritmos definidos como parte del proceso de fusión son:

- Inferencia Bayesiana. Este algoritmo asume una representación de un estado de interés dada una observación, es decir, la relación entre el estado y la observación pueden ser codificados como una probabilidad conjunta o una distribución de probabilidad conjunta [32]. De esta forma la información es actualiza cada vez que se fusionan nuevos datos derivados de una nueva observación [31]. Esto provee una forma de combinar datos observados con información previamente adquirida [37].
- Vecinos cercanos. Consiste en agrupar o seleccionar los datos cercanos unos con otros. La cercanía de un valor depende de la métrica de la distancia utilizada, por ejemplo basada en la distancia absoluta, euclidiana o una función estadística de distancia [15, 31].
- Filtro de Kalman. Este algoritmo utiliza una serie de estimaciones de estados previos para hacer una predicción de estados siguientes, adelantando las decisiones del sistema [7, 31, 37], esto es, hace una estimación del estado que evoluciona con el tiempo basándose en observaciones periódicas. Para esto el filtro de Kalman emplea un modelo estadístico sobre la evolución de un parámetro contra el tiempo y cómo las observaciones se relacionan con este parámetro.

Para la *validación* de la fusión de datos se incluye como métodos de evaluación la simulación de Monte Carlo y las técnicas de análisis de error de covarianza.

- Simulación de Monte Carlo. Se encarga de realizar un análisis de la fusión de datos basado en la creación de modelos de posibles resultados, esto es, mediante la sustitución de un rango de valores para un determinado valor con incertidumbre [7]. Una característica de este tipo de simulación es el uso de mediciones estadísticas para medir el desempeño del proceso de fusión [37].
- Análisis de error de covarianza. Es un modelo lineal general con una o más variables cuantitativas y uno o más factores. Se calcula mediante una regresión

lineal múltiple, que es un procedimiento estadístico que elimina la heterogeneidad causada en la variable de interés (dependiente) por la influencia de una o más variables cuantitativas (covariables) [15]. La inclusión de covariables puede aumentar la potencia estadística debido a que por lo general reduce la variabilidad.

Finalmente, en la etapa de *carga* se gestiona el almacenamiento de los datos ya sea en un repositorio, almacén, archivo, u otros medios de acuerdo a las necesidades de uso (texto, símbolo, características, decisiones o patrones), como la analítica de datos para detectar patrones de comportamiento del usuario, roles, perfiles, emociones, configuraciones, trabajo colaborativo, y otros temas de interés asociados a los sistemas consientes del contexto; dominio que ha demostrado ser sumamente útil para percibir, capturar y analizar datos de diversos ámbitos, como sociales, comerciales, biomédicos, ambientales, militares, entre otros.

## **5. Consideraciones finales**

La fusión de datos en la actualidad ha tomado una creciente importancia debido a la incorporación de sensores en los sistemas de cómputo y a la necesidad de hacer que estos sistemas interactúen con el usuario y el ambiente, haciendo que éstos se adapten a situaciones que ocurren en el entorno y den el servicio adecuado en el momento en que se requiera.

La principal ventaja de fusionar datos, en comparación de tener una sola fuente, es tener mayor información para caracterizar una determinada situación o actividad que es observada. Sin embargo, para tener diferentes fuentes es necesario usar diversos sensores que pueden variar en naturaleza, formato, velocidad e incluso interactuar con fuentes lógicas, como bases de datos, bitácoras o preferencias del usuario.

La heterogeneidad de datos trae consigo retos que deben ser solucionados, donde la imperfección en los datos, el dinamismo del entorno y la propia naturaleza de éstos crean inconsistencias que deben resolverse mediante métodos adecuados.

Resolver el reto de la inconsistencia de datos en el proceso de fusión no es trivial debido a la amplia variedad de datos heterogéneos que se producen a través de sensores, objetos y aplicaciones. Los cuales son útiles para el análisis del contexto y la caracterización de una determinada situación.

Derivado de la primera etapa del trabajo de investigación sobre la fusión de datos de fuentes heterogéneas, se propuso el diseño conceptual de un método de fusión de datos estructurado en cuatro etapas: a) extracción, b) pre-procesamiento, c) fusión y d) carga. El enfoque de esta propuesta es detectar la inconsistencia de los datos, previo al proceso de fusión, para mantener su consistencia durante la integración, evitando el conflicto, el desorden y el error en los datos.

Cómo trabajo posterior se implementará el método de fusión de datos y se evaluará el desempeño de los algoritmos desde el punto de vista de cumplimiento del objetivo, que es de mantener la consistencia de los datos en el proceso de fusión. Para esto se utilizarán datos heterogéneos obtenidos de un sistema consiente del contexto y la actividad del usuario.

Finalmente, definir y diseñar nuevos métodos de fusión de datos representa en la actualidad un reto importante para mejorar la precisión de los sistemas de integración de datos, que pueden ser generados por las interacciones del usuario con el sistema y el ambiente. Esto representa un campo de oportunidad para la investigación de sistemas conscientes del contexto y la interacción humano–computadora con el propósito de mejorar el trabajo colaborativo entre usuarios, optimizar las tareas realizadas, dar soporte al proceso de toma de decisiones, realizar inferencias, entre otros.

**Agradecimientos.** Este trabajo forma parte del proyecto “Infraestructura para agilizar el desarrollo de sistemas centrados en el usuario” financiado por el Consejo Nacional de Ciencia y Tecnología (CONACYT) en el marco del programa de Cátedras CONACYT (Ref. 3053). Además, se agradece a CONACYT por la beca de doctorado número 449416, así como a la Universidad Veracruzana por el apoyo en el desarrollo de esta investigación.

## Referencias

1. Gellersen, H., Schmidt, A., Beigl, M.: Multi–sensor context–awareness in mobile devices and smart artifacts. *Mobile Networks and Applications*, 7(5), pp. 341–351 (2002)
2. Benítez–Guerrero, E.: Context–Aware Mobile Information Systems: Data Management Issues and Opportunities. In: *Proceedings of the International Conference on Information & Knowledge Engineering*, Las Vegas (2010)
3. Wu–Zhu, X., Wu, G., Ding, W.: Data mining with big data. *IEEE transactions on knowledge and data engineering*, 26(1), pp. 97–107 (2014)
4. Abowd, G. D., Mynatt, E. D.: Charting past, present, and future research in ubiquitous computing. *ACM Transactions on Computer Human Interaction*, 7(1), pp. 29–58 (2000)
5. Bernardos, A. M., Tarrío, P., Casar, J. R.: A data fusion framework for context–aware mobile services. In: *Proceedings Multisensor Fusion and Integration for Intelligent Systems*, pp. 606–613 (2008)
6. White, F.: Data fusion lexicon. Reporte técnico, Joint Directors of Laboratories (1991)
7. Hall, D., Llinas, J.: An introduction to multisensory data fusion. *Proceedings of the IEEE*, 85(1), pp. 6–23 (1997)
8. Dasarathy, B.: Sensor fusion potential exploitation–innovative architectures and illustrative applications. *Proceedings of the IEEE*, 85(1), pp. 24–38 (1997)
9. Rodríguez, S., De Paz, J., Villarrubia, G., Zato, C., Bajo, J., Corchado, J.: Multi-agent information fusion system to manage data from a WSN in residential home. *Information Fusion*, 23, pp. 43–57 (2015)
10. Surie, D., Pederson, T., Lagriffoul, F., Janlert, L., Sjölie, D.: Activity recognition using an egocentric perspective of everyday objects. In: *International conference on ubiquitous intelligence and computing*, Hong Kong, pp. 246–257 (2007)
11. Branton, C., Ullmer, N., Wiggins, A., Rogge, L., Setty, N., Beck, S. D., Reeser, A.: Toward rapid and iterative development of tangible, collaborative, distributed user interfaces. In: *Proceedings of the 5th ACM SIGCHI symposium on Engineering interactive computing systems*, London, pp. 239–248 (2013)
12. Yang, M., Al Kutubi, M., Pham, D.: Continuous acoustic source tracking for tangible acoustic interfaces. *Measurement*, 46(3), pp. 1272–1278 (2013)

13. Jenkins, M. P., Gross, G. A., Bisantz, A. M., Nagi, R.: Towards context aware data fusion: Modeling and integration of situationally qualified human observations to manage uncertainty in a hard+soft fusion process. *Information Fusion*, 21, pp.130–144 (2015)
14. Qin, W., Suo, Y., Shi, Y.: CAMPS: A middleware for providing context-aware services for smart space. In: *International conference on grid and pervasive computing*, Taichung, pp. 644–653 (2006)
15. Castanedo, F.: A review of data fusion techniques. *The Scientific World Journal* (2013)
16. Von-Neumann, J.: Probabilistic logics and the synthesis of reliable organisms from unreliable components. *Automata studies*, 34, pp. 43–98 (1956)
17. Chow, C.: Statistical independence and threshold functions. *IEEE Transactions on Electronic Computers*, 1, pp. 66–68 (1965)
18. Hashem, S.: Algorithms for optimal linear combinations of neural networks. *Neural networks*, 1, pp. 242–247 (1997)
19. Varshney, P. K.: Distributed Bayesian detection: Parallel fusion network. In: *Distributed detection and data fusion*, Springer, New York, pp. 36–118 (1997)
20. Breiman, L.: Stacked regressions. *Machine learning*, 24(1), pp. 49–64 (1996)
21. Juditsky, A., Nemirovski, A.: Functional aggregation for nonparametric estimation. In: *IRISA, Rennes* (1996)
22. Granger, C. W.: Invited review-combining forecasts twenty years later. *Journal of Forecasting*, 8(3), pp. 167–173 (1989)
23. Jagadish, H., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., Shahabi, C.: Big data and its technical challenges. *Communications of the ACM*, 57(7), pp. 86–94 (2014)
24. Luo, R., Kay, M.: Multisensor integration and fusion: Issues and approaches. In: *Proceedings SPIE, Florida*, pp. 42–49 (1988)
25. Luo, R., Yih, C., Su, K.: Multisensor fusion and integration: approaches, applications, and future research directions. *IEEE Sensors Journal*, 2(2), pp. 107–119 (2002)
26. Thomopoulos, S. C.: Sensor integration and data fusion. In: *Symposium on visual communications, image processing, and intelligent robotics systems*, pp. 178–191 (1990)
27. Luo, R. C., Kay, M. G.: Multisensor integration and fusion: issues and approaches. In: *Orlando technical symposium*, pp. 42–49 (1988)
28. Pau, L. F.: Sensor data fusion. *Journal of Intelligent and Robotic Systems*, 1(2), pp. 103–116 (1988)
29. Harris, C., Bailey, A., Dodd, T.: Multi-sensor data fusion in defence and aerospace. *Aeronautical Journal*, 102(1015), pp. 229–244 (1998)
30. Bedworth, M., O'Brien, J.: The omnibus model: a new model of data fusion? *IEEE Aerospace and Electronic Systems Magazine*, 15(4), pp. 30–36 (2000)
31. Khaleghi, B., Khamis, A., Karray, F., Razavi, S.: Multisensor data fusion: A review of the state-of-the-art. *Information Fusion*, 14(1), pp. 28–44 (2013)
32. Kumar, M., Garg, D. P., Zachery, R. A.: A method for judicious fusion of inconsistent multiple sensor data. *IEEE Sensors Journal*, 7(5), pp. 723–733 (2007)
33. Kale, D. R., Aparadh, S. Y.: A Study of a Detection and Elimination of Data Inconsistency in Data Integration. *International Journal of Scientific Research in Science, Engineering and Technology*, 2(1), pp. 532–535 (2016)
34. Saha, B., Srivastava, D.: Data quality: The other face of big data. In: *IEEE 30th International Conference on Data Engineering*, pp. 1294–1297 (2014)
35. Wang, X., Huang, L., Xu, X., Zhang, Y., Chen, J. Q.: A Solution for Data Inconsistency in Data Integration. *Journal of Information Science and Engineering* 27(2), pp. 681–695 (2011)

*Julio Muñoz, Guillermo Molero–Castillo, Edgard Benítez–Guerrero*

36. Patten, J., Ishii, H., Hines, J., Pangaro, G.: Sensetable: a wireless object tracking platform for tangible user interfaces. In: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 253–260 (2001)
37. Siciliano, B., Khatib, O.: Handbook of Robotics. Springer (2016)