

**Centro de Investigación Científica y de Educación
Superior de Ensenada, Baja California**



**Doctorado en Ciencias
en Ciencias de la Computación**

**Detección y visualización de comportamientos no
usuales en adultos mayores institucionalizados a
través de un análisis híbrido**

Tesis

para cubrir parcialmente los requisitos necesarios para obtener el grado de
Doctor en Ciencias

Presenta:

Valeria Soto Mendoza

Ensenada, Baja California, México

2017

Tesis defendida por

Valeria Soto Mendoza

y aprobada por el siguiente Comité

Dr. José Antonio García Macías
Director de tesis

Dra. Ana Isabel Martínez García

Dr. Hugo Homero Hidalgo Silva

Dr. Rolando Menchaca Méndez



Dr. Jesús Favela Vara
Coordinador del Posgrado en Ciencias de la Computación

Dra. Rufina Hernández Martínez
Director de Estudios de Posgrado

Valeria Soto Mendoza © 2017

Queda prohibida la reproducción parcial o total de esta obra sin el permiso formal y explícito del autor y director de la tesis

Resumen de la tesis que presenta Valeria Soto Mendoza como requisito parcial para la obtención del grado de Doctor en Ciencias en Ciencias de la Computación

Detección y visualización de comportamientos no usuales en adultos mayores institucionalizados a través de un análisis híbrido

Resumen aprobado por:

Dr. José Antonio García Macías

Director de tesis

En las residencias geriátricas, los cuidadores de adultos mayores llevan bitácoras donde registran las actividades diarias mientras atienden a los residentes. Para este trabajo se instrumentó uno de estos lugares con infraestructura física de sensado para el seguimiento continuo de los adultos mayores. Toda la información generada, tanto la colectada por cuidadores como la proveniente de los sensores, se almacena en repositorios de actividades. El continuo registro de las actividades de adultos mayores es relevante porque a partir de las actividades es posible observar indicadores sobre su estado de salud (por ejemplo: deterioros físicos o mentales, depresión, entre otros). Es importante también observar los cambios en las actividades a través del tiempo, utilizando información histórica. De esta forma se puede estudiar el comportamiento. En computación el comportamiento se ha estudiado desde dos perspectivas distintas: una guiada por datos y otra guiada por conocimiento. En este trabajo se propone un componente híbrido que involucra ambas perspectivas con el fin de analizar el comportamiento y detectar anomalías en las actividades del adulto mayor. El componente híbrido fue probado con datos reales de 45 adultos mayores correspondientes a 39 meses de colección de datos. Como resultado se proponen tres metáforas para la visualización de datos que resultaron ser adecuadas para observar anomalías en las actividades de adultos mayores.

Palabras clave: Análisis de comportamientos anormales, adultos mayores, residencias geriátricas.

Abstract of the thesis presented by Valeria Soto Mendoza as a partial requirement to obtain the Doctor of Science degree in Computer Science

Detection and visualization of unusual behaviors from institutionalized older adults through a hybrid analysis

Abstract approved by:

Dr. José Antonio García Macías
Thesis Director

In geriatric centers or nursing homes, caregivers of older adults record daily activities while caring for residents. For this work, we equipped one of these centers with sensing infrastructure to monitor older adults continuously. All the information generated is stored in activity repositories. Recording activities of older adults is relevant because from the records it is possible to observe indicators about their health status (for example, physical or mental deterioration, depression, among others). It is also important to observe changes in activities over time, using historical information. This enables the study of behavior. In computer sciences the behavior has been studied from two different perspectives: one driven by data and another one driven by knowledge. In this thesis, a hybrid component is proposed that involves both perspectives in order to analyze the behavior and detect abnormalities in the activities of the older adult. The hybrid component was tested with real data from 45 older adults corresponding to 39 months of data collection. As a result, three metaphors are proposed for the visualization of data which proved to be adequate to observe abnormalities in the activities of older adults.

Keywords: Abnormal behaviour analysis, Older adults, Geriatric residences.

Dedicatoria

A todos los que me acompañaron de cerca en este proceso

A mi madre, padre y hermanos que me apoyaron moralmente en la distancia

Agradecimientos

Al Dr. José Antonio García Macías por su tiempo, paciencia y orientación a lo largo de este trabajo de tesis.

Al Dr. Edgar Leonel Chávez González por ser un mentor y asesor académico, e incluso un amigo; sus reflexiones y guía siempre fueron interesantes y muy relevantes para el trabajo. Gracias por la disposición y el tiempo invertido.

A los miembros del comité: la Dra. Ana Isabel Martínez García, el Dr. Hugo Homero Hidalgo Silva y el Dr. Rolando Menchaca Méndez por sus comentarios y observaciones que enriquecieron este trabajo. Dra. Ana por brindarme su amistad y guiarme en los momentos complicados.

A la Dra. Patricia Serrano Alvarado por su disposición a colaborar durante mi estancia en Nantes, Francia. Sus consejos fueron más allá que puros consejos relacionados con aspectos académicos.

Al Dr. Rolando y Dr. Ricardo Menchaca Méndez por su apoyo y colaboración durante la estancia en el CIC-IPN, gracias por "hacerme el paro".

Al Dr. Jesús Favela Vara que siempre estuvo dispuesto a ayudar, sus comentarios y observaciones siempre fueron propositivos.

A los cuidadores, enfermeros y estudiantes participantes en las distintas evaluaciones. Al personal de Serena: Miguel Torres, Flavio Olivieri y Carlos Román por las facilidades brindadas; a Rodolfo Hernández de Arkus por proporcionar los datos. A Argel Grisolle de Acam Residencial por sus atenciones y disponibilidad a colaborar en este proyecto.

Al Dr. Héctor Efraín Ruiz y Ruiz por aguantarme durante la escritura de la tesis y la parte final de este recorrido.

A mis amigos (casi hermanos) que me acompañaron en este largo camino, mis incondicionales que adoro Víctor, Kary, Jess, Maythé, Cathy, Jehú, Marco y Héctor.

A Eduardo, Tona y Omar porque las desveladas fueron más llevaderas con su acompañamiento y colaboración.

A mi *roomie* Mario que verificaba que siguiera viva y que me brindó el apoyo constante a distancia.

A Dago y Memo que aunque fue corto tiempo de convivencia, han llegado a ser grandes colegas y amigos, gracias por su colaboración en las evaluaciones.

Al Centro de Investigación Científica y de Educación Superior de Ensenada.

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por brindarme el apoyo económico para realizar mis estudios de doctorado.

Tabla de contenido

	Página
Resumen en español	ii
Resumen en inglés	iii
Dedicatoria	iv
Agradecimientos	v
Lista de figuras	ix
Lista de tablas	x
Capítulo 1. Introducción	
1.1. Antecedentes	1
1.1.1. El envejecimiento	1
1.1.2. La vida en los centros de cuidados geriátricos	2
1.2. Planteamiento del problema	2
1.2.1. Preguntas de investigación	3
1.2.2. Objetivos	4
1.2.2.1. Objetivo general	4
1.2.2.2. Objetivos específicos	4
1.3. Metodología	5
1.4. Estructura de la tesis	6
Capítulo 2. Análisis de comportamiento	
2.1. Definición de comportamiento	8
2.2. Estudio del comportamiento	9
2.2.1. Enfoque guiado por datos (<i>DDA, Data-driven approach</i>)	9
2.2.2. Enfoque guiado por conocimiento (<i>KDA, Knowledge-driven approach</i>)	10
2.2.3. Enfoque híbrido (<i>HDA, Hybrid-driven approach</i>)	11
2.3. Detección de anomalías	11
2.4. Resumen de trabajo previo	12
Capítulo 3. Descripción del repositorio de datos	
3.1. Tipos de datos	14
3.2. Análisis del repositorio de actividades	15
3.3. Estructura de datos	17
3.4. Campañas de sensado	17
3.5. Resumen	20
Capítulo 4. Componente híbrido para el análisis del registro de actividad	
4.1. Descripción general	22
4.2. Pre-procesado	23

Tabla de contenido (continuación)

4.3.	Análisis	26
4.3.1.	Análisis basado en datos	27
4.3.2.	Análisis basado en conocimiento	30
4.4.	Predicción	32
4.4.1.	Detección de anomalías basadas en tiempo.	32
4.4.2.	Detección de anomalías basadas en secuencia.	33
4.4.3.	Detección de anomalías con inferencias semánticas	34
4.5.	Visualización	34
4.5.1.	Pirámide de actividades	35
4.5.2.	Mapas diarios de colores	35
4.5.3.	Línea de actividades	37
4.6.	Resumen	37
Capítulo 5. Experimentos y resultados		
5.1.	Detección automática de anomalías	39
5.2.	Evaluación humana de anomalías	46
5.3.	Validación del componente híbrido	49
5.4.	Otras alternativas de desarrollo	53
5.4.1.	Sesión de evaluación del diseño.	58
5.5.	Resumen	62
Capítulo 6. Conclusiones, aportaciones y trabajo futuro		
6.1.	Conclusiones	63
6.2.	Aportaciones	65
6.3.	Limitaciones	68
6.4.	Trabajo futuro	69
Literatura citada		73
Anexos		80

Lista de figuras

Figura	Página
1. Metodología de investigación.	5
2. Comparación entre perspectivas de estudio del comportamiento.	9
3. Estructura general del registro de actividad.	17
4. Diagrama del componente híbrido para el análisis del registro de actividad.	23
5. Análisis del tiempo de ocurrencia de cada actividad.	28
6. Análisis de las secuencias de actividades frecuentes.	29
7. Compresión de la secuencia de actividades.	30
8. Representación semántica del registro de actividad.	31
9. Diagrama general de <i>ABIOAOntology</i>	32
10. Herramienta de visualización denominada pirámide de actividades.	36
11. Herramienta de visualización denominada mapas diarios de colores.	36
12. Herramienta de visualización denominada línea de actividades.	38
13. Representación gráfica de las horas a lo largo del día.	40
14. Distribución de las actividades de alimentación (<i>Feeding</i>) a lo largo del día.	41
15. Distribución de las actividades de signos vitales (<i>Vitals</i>) a lo largo del día. .	41
16. Identificación de horarios habituales de las actividades de alimentación. . .	42
17. Identificación de horarios habituales de las actividades de signos vitales. .	42
19. Frecuencia de patrones secuenciales del Residente06 con diferente longi- tud.	44
23. Matriz de transición de estados del residente S7.	55
24. Número de transiciones de estado de cada residente durante la mañana, tarde y noche.	55
25. Ejemplo de una secuencia de estados siguiendo el camino más probable extraído de la matriz de transiciones.	57
26. Camino individual siguiendo las transiciones entre estados.	58
27. Ejemplo de los registros de actividades dentro de una guardería de niños. .	71

Lista de tablas

Tabla	Página
1.	Resumen de la revisión de literatura. 13
2.	Tipos de datos contenidos en el repositorio. 15
3.	Descripción detallada de la estructura del registro de actividad. 18
4.	Características generales del repositorio de datos obtenido en una residencia geriátrica. 18
5.	Características demográficas y de salud de los participantes. 19
6.	Características del repositorio por cada participante (CS1 = campaña de sensado 1 y CS2 = campaña de sensado 2 respectivamente. 20
7.	Etapas del pre-procesado. 25
8.	Número de días y registros utilizados como datos de entrenamiento y prueba por residente. 26
9.	Características del repositorio de datos pre-procesado. 26
10.	Estadísticas obtenidas a partir del <i>TDS</i> 27
11.	Niveles de anormalidad definidos para el análisis basado en tiempo. 33
12.	Niveles de anormalidad definidos para el análisis basado en secuencia. 34
13.	Detección automática de anomalías. 46
14.	Resultados de la evaluación de metáforas con estudiantes. 47
15.	Resultados de la evaluación de metáforas con cuidadores. 48
16.	Medidas de precisión y recuerdo de la evaluación humana de anomalías. 48
19.	Lista de estados de un residente dentro de la residencia geriátrica. 53
20.	Datos demográficos de los participantes en la sesión de diseño del calendario predictivo. 59
21.	Resultados del análisis de la aplicación del TAM. 61

Capítulo 1. Introducción

1.1. Antecedentes

En años recientes a nivel global se ha incrementado el número de lugares para la atención y cuidado del adulto mayor, obedeciendo al incremento en la demanda de este tipo de servicios (Entrepreneur, 2005); asimismo, como consecuencia del aumento de la esperanza de vida y de la disminución de la tasa de fecundidad, la proporción de personas mayores de 60 años está aumentando vertiginosamente a nivel mundial (WHO, 2015). Los centros de cuidado de adultos mayores, tradicionalmente conocidos como asilos (despectivamente), ahora son centros de cuidados geriátricos especializados que se han convertido en un negocio lucrativo en pleno auge con muy buenas perspectivas de crecimiento (Stonefield, 2013). Según la Organización de las Naciones Unidas (ONU, 2015), hay casi 700 millones de personas mayores de 60 años a nivel mundial. Para 2050, las personas de 60 años o más serán 2,000 millones, esto representará más del 20% de la población mundial.

1.1.1. El envejecimiento

El envejecimiento es un proceso de cambios a través del tiempo (Robledo *et al.*, 2012). Dicho proceso se da de manera natural, gradual, continua, irreversible y completa. Estos cambios se dan a nivel biológico, psicológico y social, y están determinados por la historia, la cultura y la situación económica, de los grupos y las personas. Cada individuo envejece de manera diferente, dependiendo de sus características innatas, de las que adquiere a través de la experiencia y de las circunstancias a las que se haya enfrentado durante su vida; todo esto contribuye a que tenga un nivel de dependencia asociado con la forma en la que envejeció. Los adultos mayores con un nivel de dependencia alto frecuentemente son cuidados por familiares o amigos cercanos, es decir, *cuidadores informales*. En otros casos, en donde se cuentan con los recursos económicos suficientes, son trasladados a vivir en lugares dedicados exclusivamente para el cuidado de adultos mayores, conocidos como *residencias geriátricas*. En donde viven bajo la supervisión de personal médico y cuidadores capacitados (*cuidadores formales*), (Robledo *et al.*, 2015).

1.1.2. La vida en los centros de cuidados geriátricos

Con base en estudios empíricos realizados previamente (Soto Mendoza, 2012), se ha observado que en las residencias geriátricas se llevan bitácoras de las actividades que se realizan día con día (Archer *et al.*, 2011). Recientemente se han desarrollado herramientas tecnológicas que permiten que los cuidadores registren las actividades por medios electrónicos mientras atienden a los residentes (Kane y Wilson, 2007; Zúñiga Rojas, 2012). La información generada se almacena en repositorios de actividades. Estos repositorios contienen información sobre las actividades del adulto mayor, así como también sobre mediciones obtenidas a través de instrumentos de captura automática (por ejemplo, la temperatura, los signos vitales, el peso, etc.). Adicionalmente, el cuidador puede enriquecer los datos anteriores con información basada en su propia experiencia y conocimiento acerca del residente. Las anotaciones de los cuidadores son más certeras y brindan más contexto que las predicciones que realice cualquier técnica de reconocimiento de actividades actualmente.

1.2. Planteamiento del problema

El registro de las actividades de adultos mayores es relevante porque a partir de ellas es posible observar indicadores sobre su estado de salud (Merilahti *et al.*, 2012; Chernbumroong *et al.*, 2013) (por ejemplo: deterioros físicos o mentales, depresión, entre otros). Es importante también observar los cambios en las actividades a través del tiempo, utilizando información histórica. De esta forma se puede estudiar el comportamiento (Fernández-Llatas *et al.*, 2011), el cual es definido ambiguamente como un conjunto de actitudes, estados de ánimo, hábitos alimenticios, entre otros.

Los cuidadores en centros geriátricos, además de realizar las actividades de cuidado y atención del adulto mayor, efectúan también otras actividades administrativas en donde se dan el tiempo de realizar el registro de actividades y anotaciones adicionales sobre el estado del adulto mayor. En el mejor de los casos, el registro se realiza en tiempo real; sin embargo, en otras ocasiones la excesiva carga de trabajo impide que lleven un registro actualizado. Sin mencionar el hecho de hacer consultas de información; las cuales se realizan únicamente cuando algún acontecimiento extraordinario ocurre (por ejemplo, cuando un adulto mayor lleva varios días sin obrar

o sin bañarse). En este caso los cuidadores y el personal médico recurren a revisar los registros históricos buscando los datos relevantes. La búsqueda y discriminación de información puede consumirles un tiempo considerable, que los puede distraer de otras actividades de atención prioritarias.

Algunas investigaciones han propuesto herramientas de asistencia para el cuidado y seguimiento de la salud en adultos mayores (Daniel *et al.*, 2009; Joshi *et al.*, 2013). Sin embargo, son escasas las tecnologías de asistencia enfocadas en facilitar la labor de los cuidadores de adultos mayores, mientras que las herramientas que existen están enfocadas en mitigar los efectos físicos y mentales de la carga de trabajo en el cuidador (Corno *et al.*, 2016). Lo anterior denota una brecha de oportunidad para proporcionar herramientas que apoyen la labor de los cuidadores de adultos mayores.

Resumiendo, a pesar de que se cuentan con registros de actividades históricos no se consultan con frecuencia. Tampoco se cuenta con herramientas que de forma automática emitan valoraciones, factoricen y extraigan la información relevante para que sea consultada por los cuidadores y los encargados de la salud del adulto mayor. Es por esto que en esta investigación se proponen herramientas y técnicas que de manera automática analicen los registros de actividades, con la finalidad de extraer información relevante. Asimismo, se proponen metáforas de visualización para los datos simplificados que permitan a los cuidadores observar los cambios de comportamiento en el adulto mayor.

1.2.1. Preguntas de investigación

Con fundamento en la problemática anterior, se plantean las siguientes preguntas de investigación:

[PI1] ¿Cuál debe ser la forma de representar los datos del repositorio de actividades de un adulto mayor que vive en una residencia geriátrica para la detección de patrones?

[PI2] ¿Qué tipo de patrones de actividad se pueden encontrar mediante el análisis de las actividades de la vida diaria del adulto mayor que vive en una residencia geriátrica?

[PI3] ¿Qué tipo de técnicas de análisis se pueden emplear para detectar variaciones en las actividades del adulto mayor que vive en una residencia geriátrica?

[PI4] ¿Cuáles son las formas más eficaces, desde la perspectiva del cuidador, para visualizar los datos de actividades del adulto mayor que vive en una residencia geriátrica?

1.2.2. Objetivos

1.2.2.1. Objetivo general

Considerando las oportunidades en cuanto a la disponibilidad de tecnología de asistencia dirigida a cuidadores de adultos mayores y a la problemática presentada anteriormente; así como con el fin último de responder las preguntas de investigación, se establece el objetivo general de esta investigación como sigue:

Detección de patrones de actividad a través del análisis de un repositorio de actividades de adultos mayores institucionalizados para visualizar comportamientos no usuales relacionados a su bienestar.

1.2.2.2. Objetivos específicos

Como parte del objetivo general se plantean los siguientes objetivos específicos:

[OE1] Obtener un repositorio de datos a través de un estudio longitudinal, el cual contenga los patrones de actividades implícitos de residentes.

[OE2] Analizar el repositorio de datos obtenido para conocer las características específicas de los datos.

[OE3] Proponer una representación adecuada que incluya los distintos tipos de datos contenidos en el repositorio de datos.

[OE4] Establecer técnicas de análisis para la detección personalizada de anomalías utilizando el repositorio de datos.

[OE5] Proponer herramientas para apoyar al cuidador en la visualización de los datos y posibles desviaciones encontradas.

[OE6] Evaluar las herramientas para la visualización propuestas con cuidadores de una residencia geriátrica.

1.3. Metodología

La metodología seguida a lo largo de este trabajo de tesis se describe a continuación (ver Figura 1):

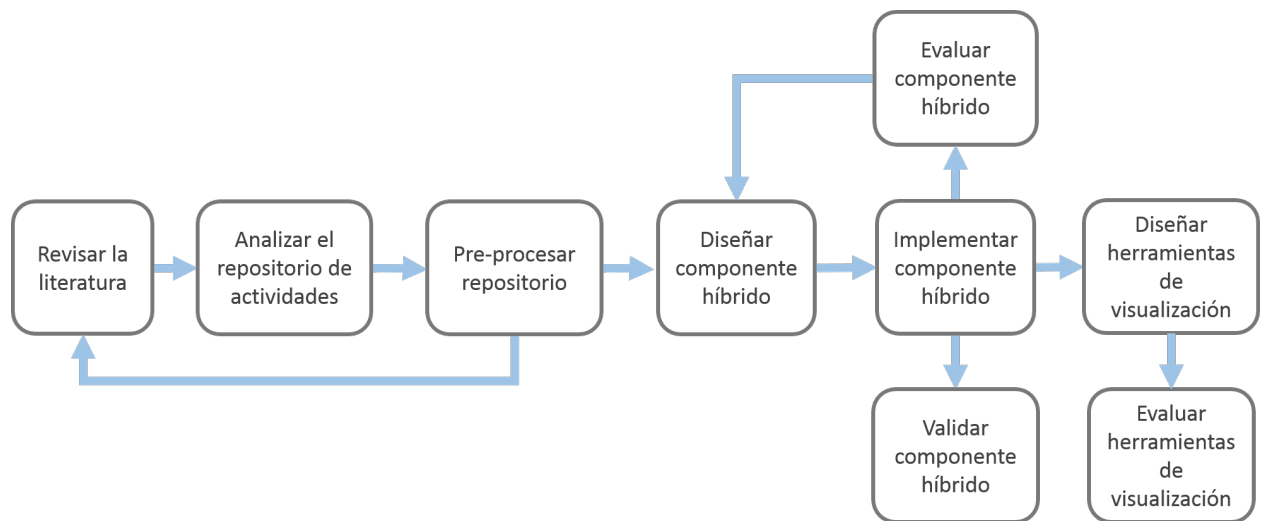


Figura 1. Metodología de investigación.

Revisión de la literatura. Se realizó una revisión exhaustiva de la literatura para establecer el estado del arte sobre análisis de comportamiento. Definiendo el concepto de comportamiento, conociendo las técnicas y las representaciones empleadas para su análisis.

Análisis del repositorio de actividades. Se realizó un análisis preliminar de un repositorio obtenido de una residencia geriátrica para familiarizarse con los datos, conocer las características de los mismos y definir las técnicas de análisis más apropiadas.

Pre-procesamiento del repositorio de actividades. Este proceso fue realizado para homogenizar la base de datos, detectar errores en los datos y corregirlos.

Las siguientes tres fases fueron parte de un proceso iterativo en donde se determinaron las técnicas de análisis más adecuadas con base al repositorio de datos con el que se contaba.

Diseño del componente híbrido. Con base en los aspectos detectados en la fase de análisis del repositorio de actividades, se propone un componente híbrido de análisis del registro de actividad. El diseño de este componente involucra aspectos de técnicas basadas en datos y de técnicas basadas en conocimiento.

Implementación del componente híbrido. La implementación del componente siguiendo el enfoque basado en datos utilizó técnicas como clúster, minería de patrones secuenciales, subsecuencia común más larga y emparejamiento recursivo. En cuanto al enfoque basado en conocimiento, se implementó una ontología y un conjunto de reglas semánticas.

Evaluación del componente híbrido. Una vez que se implementó el componente híbrido, iterativamente se realizaron experimentos para refinar las técnicas empleadas.

Validación del componente híbrido. Esta validación consiste en incluir casos de prueba específicos extraídos del conjunto de datos para la evaluación holística del componente.

Diseño de herramientas para la visualización. Con el objetivo de facilitar la visualización de los datos analizados por el componente híbrido, se proponen algunos diseños de metáforas para la visualización de datos para apoyar a los cuidadores de adultos mayores.

Evaluación de las herramientas de visualización. Esta fase consiste en la evaluación cualitativa de las metáforas de visualización propuestas con cuidadores de una residencia geriátrica.

1.4. Estructura de la tesis

La presente tesis está conformada por seis capítulos, los cuales se describen a continuación: en el capítulo dos, se presenta el trabajo relacionado al análisis de comportamiento abarcando desde la definición de comportamiento, los enfoques en que

se estudia el comportamiento y la detección de anomalías.

En el capítulo tres se describe el conjunto de datos empleado en esta investigación, las características generales de los datos y de cada tipo de dato. Así como un análisis preliminar de este repositorio de datos.

En el capítulo cuatro se desarrolla el componente híbrido para el análisis del comportamiento de adultos mayores. Se detallan los elementos que lo conforman, así como su funcionamiento.

Los experimentos y discusión de resultados se describen en el capítulo cinco.

Por último, en el capítulo seis se presentan las conclusiones, aportaciones, limitaciones y trabajo futuro propuesto derivado de este trabajo de tesis.

Capítulo 2. Análisis de comportamiento

En este capítulo se desarrolla el marco teórico conceptual que sirve de base para el desarrollo de este trabajo de tesis. Se establecen los tópicos y conceptos sobre los que se sustenta la presente investigación, los cuales ayudarán al lector en el entendimiento del documento de tesis. Primeramente, se aborda el concepto de comportamiento, para posteriormente abrir paso a las técnicas que se han empleado para el modelado y análisis del mismo. Se concluye el capítulo con un breve resumen del trabajo previo relevante.

2.1. Definición de comportamiento

El comportamiento es un concepto que ha cobrado importancia en diferentes áreas: científica, social, económica, cultural, política, militar, entre otras. La Asociación Estadounidense de Psicología (APA, por sus siglas en inglés American Psychological Association) define comportamiento como *las acciones mediante las cuales un organismo se ajusta a su entorno* (Gerrig et al., 2015). En las ciencias sociales se define como *la respuesta del organismo a varios estímulos* (Minton, 2013). Las ciencias naturales lo definen como *el conjunto de respuestas coordinadas internamente de los organismos vivos hacia estímulos internos y/o externos* (Levitis et al., 2009) o *cualquier acción de un organismo que cambia su relación con su entorno* (Dusenbery, 2009). En el diccionario el comportamiento está definido como *las acciones o reacciones de cualquier organismo o material en respuesta a un estímulo interno o externo bajo ciertas condiciones* (Oxford, 2017).

En las ciencias de la computación, el comportamiento no está definido de manera estándar. Las metodologías, técnicas y herramientas prácticas para representar, modelar, analizar, entender y utilizar comportamientos (humanos, organizacionales, sociales, artificiales y virtuales), interacciones y relaciones entre comportamientos, redes de comportamiento, patrones de comportamiento, impacto del comportamiento, la conformación y descomposición de grupos orientados por comportamientos, inteligencia colectiva, así como el surgimiento de inteligencia del comportamiento, son aspectos que estudia el área conocida como *behavior computing* o *behavior informatics* (Cao y Philip, 2012). Existen también otros campos como *behavior-aware computing*

que surgen para el estudio del comportamiento (Favela, 2013).

Para el ámbito de este trabajo de tesis, y tomando como referencia los trabajos previos de Kim *et al.* (2010); Aztiria *et al.* (2012); Rashidi y Cook (2013) y Lago *et al.* (2015), el comportamiento se define como ***una secuencia de actividades repetitiva, tomando en cuenta la frecuencia de la secuencia de actividades dentro de un periodo de tiempo específico.***

2.2. Estudio del comportamiento

En la revisión de literatura realizada, se han identificado varias perspectivas para estudiar el comportamiento. En la Figura 2 se muestra una comparativa entre los distintos enfoques. A continuación se explican los detalles de cada enfoque.

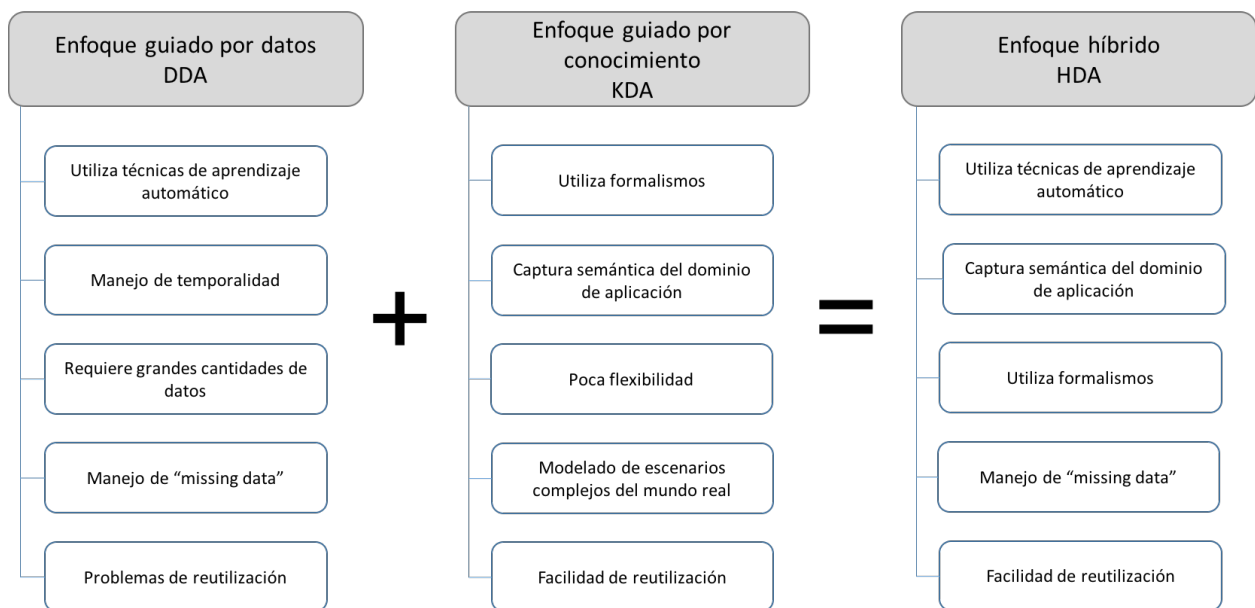


Figura 2. Comparación entre perspectivas de estudio del comportamiento.

2.2.1. Enfoque guiado por datos (DDA, *Data-driven approach*)

Este enfoque utiliza técnicas de aprendizaje automático para capturar las dependencias temporales que permiten el manejo de información con respecto al tiempo (Rodríguez *et al.*, 2014). Dentro de las desventajas de esta perspectiva es que se requieren grandes cantidades de datos para el entrenamiento de los modelos, los cuales

pueden ser difíciles de conseguir. Asimismo, es difícil la reutilización de los modelos porque son datos correspondientes a un usuario y cada usuario realiza sus actividades de forma diferente (y esas diferencias no se capturan). La ventaja de estas técnicas orientadas a los datos es que pueden trabajar con datos faltantes (*missing data*) o se pueden determinar con técnicas estadísticas (Fatima *et al.*, 2013).

Dawadi *et al.* (2016) proponen un análisis de actividades para detectar cambios de comportamiento en rutinas basadas en una secuencia de actividades; la duración de la actividad se utiliza para comparar entre diferentes curvas de actividad. Este enfoque es útil cuando las actividades diarias son reconocidas automáticamente a partir de datos sin procesar usando algoritmos de reconocimiento de actividad. En donde el estado de un sensor impacta en la actividad que se está ejecutando y por lo tanto reconociendo. Sin embargo, ejecutar un análisis basado únicamente en la duración de una actividad presenta una desventaja cuando las actividades son anotadas por los cuidadores, ya que éstos simplemente indican el momento en que ocurrió la actividad. Por otro lado, Rieping *et al.* (2014) proponen un análisis del comportamiento basado en la ocupación de las áreas dentro de la casa, recolectados a través de una red de sensores. En lugar de realizar sólo un análisis cuantitativo, los autores también realizan un análisis cualitativo para comparar los resultados con las anotaciones originales. Esto sugiere que se necesita más información no sólo aquella que se origina a partir de los sensores, con el objetivo de incorporar mayor semántica en el contexto y enriquecer los resultados del proceso de análisis cuantitativo o guiado por datos.

2.2.2. Enfoque guiado por conocimiento (KDA, Knowledge-driven approach)

En este enfoque se utilizan formalismos lógicos para capturar la semántica del dominio de aplicación y crear modelos explícitos (Gómez A *et al.*, 2014). Es posible incorporar reglas basadas en el sentido común y heurísticas. Una de sus ventajas es que es fácil reutilizarlas ya que se captura la información relacionada al dominio de aplicación; sin embargo, se debe capturar todo el dominio porque posteriormente es difícil modificar los modelos semánticos (Li *et al.*, 2010; Zhang *et al.*, 2011). Para aprovechar las ventajas de ambas perspectivas, se han creado también modelos híbridos para el análisis de comportamiento.

2.2.3. Enfoque híbrido (*HDA, Hybrid-driven approach*)

Un enfoque híbrido combina técnicas de las perspectivas basadas en datos y las basadas en conocimiento (Okeyo *et al.*, 2014). De esta manera codifica conocimiento del dominio de aplicación y utiliza modelos bien conocidos y desarrollados de aprendizaje automático y probabilísticos. Si bien, la eficacia del sistema es independiente de los datos, está limitado por sus fuertes vínculos con la integridad de la base de conocimientos y la calidad de las reglas de inferencia (Rodríguez *et al.*, 2014). Para resarcir esta limitación, la máquina de inferencias semántica es aumentada con técnicas basadas en aprendizaje automático (Khan *et al.*, 2012a; Doucette *et al.*, 2012).

2.3. Detección de anomalías

La detección de anomalías es un objetivo final que se persigue después de analizar comportamientos normales en las actividades humanas. El descubrimiento o detección de anomalías en comportamiento se ha estudiado también a través de las perspectivas *DDA* y *KDA*. La detección de patrones anormales de vida se estudió en el trabajo de Shin *et al.* (2011), donde los autores proponen un sistema automático para el análisis de comportamiento utilizando sensores de movimiento infrarrojos para asistir a adultos mayores independientes. El flujo de datos proveniente de los sensores se analiza usando el método de descripción de datos de soporte vectorial (SVDD, support vector data description). Los resultados fueron contrastados con medidas cualitativas sobre las condiciones de salud de los adultos mayores para determinar cuándo una anomalía ocurre. Aunque los experimentos fueron realizados con datos reales obtenidos de un entorno experimental, para aplicaciones del mundo real resulta costosa la instrumentación completa del ambiente.

La propuesta de Barsocchi *et al.* (2015) es el seguimiento de adultos mayores independientes que viven solos en su casa, para detectar desviaciones en sus comportamientos a partir de las rutinas seguidas en las actividades realizadas dentro de sus casas. El análisis se realiza en un escenario real e incluye adultos mayores con alto riesgo de deterioro causado por alguna enfermedad. En lugar de utilizar redes de sensores para la adquisición de datos, los autores validaron sus predicciones comparándolas con anotaciones realizadas por observadores humanos. En este trabajo de

tesis se incorporan las anotaciones humanas para enriquecer las predicciones y el proceso de toma de decisiones.

Por otro lado, un valor atípico (en inglés conocido como *outlier*) es un dato que se desvía mucho del resto de los datos contenidos en el repositorio de datos (Patel *et al.*, 2015). La detección de valores atípicos dentro del repositorio de datos es otro método aplicado para la detección de anomalías. Mahmoud *et al.* (2012) proponen un sistema de detección de valores atípicos para las actividades de la vida diaria (ADL) dentro de una casa inteligente. Los valores atípicos pueden ser generados por un diverso número de factores (errores en los datos, valores inusuales, etc.) pero pueden ser detectados fácilmente en datos binarios debido a su sencilla representación; por lo que se requiere una representación apropiada para las anotaciones humanas. En este trabajo de tesis se propone una representación y técnicas de análisis para identificar actividades inusuales de adultos mayores con el objetivo de mejorar la calidad de cuidado personalizado.

Zhao *et al.* (2014) hacen una detección de patrones anormales en las actividades diarias usando el modelo de cadenas de Markov (MCM, Markov Chains Model). De esta manera clasifican como anormales secuencias de actividades empleando una representación espacio-temporal. Esta representación toma en cuenta la duración de cada actividad. Sin embargo, no siempre es posible calcular la duración de una actividad cuando se trata de anotaciones humanas, ya que únicamente se registra si ocurrió o no la actividad. Pero por otro lado, las anotaciones humanas pueden reducir el número de falsos positivos en la detección de anomalías causados por el mal funcionamiento de los sensores (Shin *et al.*, 2011).

2.4. Resumen de trabajo previo

En este capítulo se presentaron las perspectivas *DDA* y *KDA*. Las cuales son perspectivas que se utilizan para el análisis de comportamiento (Rodríguez *et al.*, 2014). Con la perspectiva *DDA*, se implementarán técnicas de aprendizaje automático (Witten *et al.*, 2016) y minería de datos (Candás *et al.*, 2014) para el análisis de datos (presentadas en el capítulo 4). Sin embargo, una de las desventajas de esta perspectiva es que requiere de una cantidad considerable de datos para funcionar suficientemen-

te bien. Más aún, la creación de un modelo personalizado no siempre es posible. así como su reutilización en otros dominios. Con la segunda perspectiva se utilizan formalismos para capturar la semántica de contextos complejos, de esta forma se construye un modelo semántico, el cual no es lo suficientemente flexible para admitir cambios en tiempo real (Okeyo *et al.*, 2014). Se puede concluir que ambas perspectivas se complementan en la perspectiva híbrida, *HDA*, para el análisis de comportamiento.

En la Tabla 1 se presenta un resumen del trabajo relacionado presentado en esta sección. Se detalla el tipo de análisis en el que se enfoca cada trabajo, el tipo de datos de entrada y la técnica de análisis que utilizan, y por último el parámetro que tomaron como base para el análisis.

Tabla 1. Resumen de la revisión de literatura.

Estudio	Tipo de análisis	Datos de entrada ¹	Técnica empleada	Medida de análisis
Dawadi <i>et al.</i> (2016)	Comportamiento	Actividades (S, A)	Curvas de actividad	Duración
Rieping <i>et al.</i> (2014)	Comportamiento	Ubicación (S)	LDA-Poisson-von-Mises	Evento
Shin <i>et al.</i> (2011)	Anormalidades	Movimiento (S)	SVDD	Evento
Mahmoud <i>et al.</i> (2012)	Anormalidades	Binarios (S)	SVDD	Presencia
Zhao <i>et al.</i> (2014)	Anormalidades	Ubicación (S)	MCM	Duración

En el siguiente capítulo se explica a detalle el repositorio de datos utilizado para el análisis propuesto en este trabajo de tesis. En el capítulo 4 se retomarán las perspectivas *DDA*, *KDA* y *HDA* con ejemplos concretos de aplicación, lo cual permitirá conocer mejor sus funcionamientos y alcances.

¹A = Datos colectados de anotaciones manuales. S = Datos colectados de sensores.

Capítulo 3. Descripción del repositorio de datos

En este capítulo se describe el repositorio de datos utilizado en este trabajo de tesis. A partir del análisis de los datos se identificó la fuente de donde provenían así como sus características específicas. Posteriormente, se propone una estructura de datos que unifica y representa a todos y cada uno de los tipos de datos contenidos en el repositorio.

3.1. Tipos de datos






El repositorio de datos es un conjunto de datos naturalistas, es decir, obtenidos de un entorno real, en este caso una residencia geriátrica. Una residencia geriátrica es un lugar donde viven un grupo de adultos mayores y son atendidos por uno o más cuidadores. Estos adultos mayores son denominados residentes. Desde hace varios años se ha colaborado con una residencia geriátrica ubicada en Rosarito, Baja California en el norte de México. En esta residencia viven adultos mayores con diferentes padecimientos, como por ejemplo, Alzheimer, Parkinson, deterioro físico o mental, hipertensión, diabetes, entre otros. Cada residente tiene necesidades de cuidado diferentes según su condición de salud particular, por lo que se requiere un plan de cuidados y atención personalizado.

En la residencia geriátrica se lleva una bitácora de actividades para llevar el control y seguimiento de la salud y bienestar de cada residente. En esta bitácora los cuidadores, enfermeros y personal de la residencia registran las actividades que el residente realiza diariamente. Adicionalmente, se registran eventos y situaciones fuera de la rutina. De forma tradicional las bitácoras de eventos y actividades de los residentes se llevan de forma manual, haciendo uso de lápiz y papel. Sin embargo, recientemente la residencia geriátrica ha incorporado sistemas de información para el registro digital y seguimiento de las actividades del residente (Soto-Mendoza *et al.*, 2015b). De esta forma se genera un repositorio digital de actividades de todos y cada uno de los residentes.

Este repositorio está compuesto de diferentes tipos de datos, todos relacionados al estado del residente. En el repositorio, por ejemplo, hay datos provenientes de dispositivos de sensado (con una estructura definida) y anotaciones libres (sin una estructura

definida) realizadas por los expertos en cuidado del adulto mayor. En la Tabla 2 se muestran los tipos de datos contenidos en el repositorio, así como un ejemplo de cada tipo. La *ubicación* se refiere al lugar dentro de la residencia donde se encuentra el residente. La *postura* es la posición física en que se encuentra el residente (acostado, sentado, caminando, etc.). El tipo de dato *médico* incluye medidas de signos vitales del residente adquiridas por medio de los instrumentos médicos especializados. Los *botones de pánico y dolor* son indicadores binarios de cuando el residente manifieste sentirse temeroso y/o con dolor. Las *anotaciones de cuidadores* son notas adicionales sin estructura definida que los cuidadores registran todos los días sobre la condición del residente.

Tabla 2. Tipos de datos contenidos en el repositorio.

Tipo de dato	Fuente ¹	Estructura y ejemplo
 Ubicación	S	[fecha-tiempo residente lugar] [2013-06-25T6:20:53 Resident08 Bedroom]
 Postura	S	[fecha-tiempo residente postura] [2013-10-05T15:19:52 Resident21 Walking]
 Médico	A, S	[fecha-tiempo residente temperatura presión-arterial pulso frecuencia-respiratoria glucosa peso concentración-oxígeno altura índice-masa-muscular fuerza] [2015-07-20T12:23 Resident01 36 135/78 65 18 NA 102 95 NA NA NA]
 Botones de pánico y dolor	A, S	[fecha-tiempo residente pánico dolor] [2013-02-28T18:50:37 Resident15 false true]
 Anotaciones de cuidadores	A	[fecha-tiempo cuidador residente actividad detalles] [2013-05-12T00:26 Caregiver06 Resident18 Mood {Quiet: the resident is quiet and cheerful. The flowers her family brought turned her happier.}]

3.2. Análisis del repositorio de actividades

Como primera parte de la metodología presentada en el capítulo 1 se realizó un análisis del repositorio de datos. Después de realizado el análisis se obtuvieron carac-

¹A = Datos colectados de anotaciones manuales. S = Datos colectados de sensores.

terísticas sobre los datos (Soto-Mendoza *et al.*, 2015a), las cuales son las siguientes:

Actividades heterogéneas con diferentes niveles de granularidad. Durante el día cada adulto mayor realiza diferentes actividades² que se registran a través del sistema (por ejemplo, *Vitals, Recreation, Feeding, etc.*). Además, cada actividad posee distintas descripciones adicionales que añaden valor a los datos (por ejemplo, con *Higiene* puede ser *Hands, Teeth, Clothes, etc.*). Toda esta diversidad y variabilidad en los datos hace difícil encontrar una representación única y simple de los datos (Helaoui *et al.*, 2013).

Espacios temporales. Las marcas de tiempo parecen una forma natural para establecer un orden cronológico en las actividades. Sin embargo, al observar los datos esto dista mucho de la realidad ya que en algunos días las actividades se programan a horas diferentes del día, o bien la misma actividad se realiza y se registra más de una vez en el día. Por lo tanto podrían existir espacios temporales en la secuencia que siguen las actividades (Hoque y Stankovic, 2012).

Pérdida de datos. Los registros anotados manualmente son propensos a la pérdida de datos, porque los cuidadores frecuentemente están abrumados por el trabajo, y la información pudiera no estar completa (Khan *et al.*, 2012b). En el caso de adquirir datos automáticamente, por medio de sensores, también podrían surgir problemas debido a fallos de energía o de comunicación. No se puede suponer que los registros están completos (Ohmura y Uchida, 2014).

Determinar una métrica o medida de comparación. Los datos contenidos en el repositorio están incompletos, inconsistentes y ruidosos, lo que implica un reto en diseñar un método para establecer una medida que permita identificar similitudes y desviaciones en los registros (Mahmoud *et al.*, 2011).

En general, las características listadas anteriormente se dan por: 1) los mecanismos de recolección de los datos, y 2) la semántica de las actividades. También exis-

²Los nombres de las actividades aparecerán en el idioma inglés para mantener la congruencia con el repositorio de datos que se obtuvo en este trabajo.

ten dependencias entre actividades (Soto-Mendoza *et al.*, 2015b) (es decir, *Medication* después de *Feeding*). El objetivo es explotar estas dependencias para ser incluidas en los sistemas de información y de esta manera enriquecer el proceso de cuidado del adulto mayor. Pero ¿cómo se pueden detectar automáticamente estas dependencias y relaciones? Con estas dependencias entre actividades es posible determinar cuáles relaciones son normales o anormales durante un cierto período de tiempo, ¿cómo detectarlas?; Y en caso de ser anormal, ¿cuáles son las causas de esas desviaciones? Para responder a las preguntas anteriores, primero se debe establecer una representación de los datos para posteriormente proceder al análisis de los mismos.

3.3. Estructura de datos

Los tipos de datos mostrados en la Tabla 2 se combinaron y como resultado se obtuvo una sola estructura de datos que representa a todos los datos contenidos en el repositorio. La estructura de datos se muestra en la Figura 3 y está conformada de manera general por los siguientes elementos: un número identificador, una marca de tiempo (fecha, hora y día de la semana), identificadores para cuidadores y residentes, el lugar donde se realiza la actividad, la postura del residente, el tipo de actividad y el detalle de ésta.

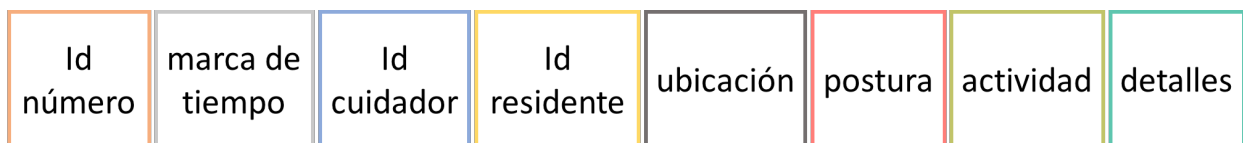


Figura 3. Estructura general del registro de actividad.

En la Tabla 3 se explica a detalle la estructura del registro de actividad presentada previamente.

3.4. Campañas de sensado

El repositorio de datos se obtuvo a través de dos campañas de sensado. La diferencia entre estas campañas de sensado es que en la campaña 1 se utilizaron los prototipos SSAMI y SMAMI propuestos por Zúñiga Rojas (2012) y Soto Mendoza (2012),

Tabla 3. Descripción detallada de la estructura del registro de actividad.

Identificador	Número del identificador único del registro de actividad.
Marca de tiempo	Fecha: formato definido de <i>Año-Mes-Día</i> .
	Tiempo: formato definido de 24 horas <i>Hora:Minuto:Segundo</i> .
Cuidador	Identificador del cuidador.
Residente	Identificador del residente o adulto mayor.
Lugar	El lugar dentro de la residencia donde se ubica el residente { <i>Bathroom, Bedroom, Courtyard, Dinning room, Garden, Gym, Inside, Kitchen, Living room, Outside, Reading room, Terrace</i> }.
Postura	La postura del residente { <i>Lay down, Fall, Sitting, Stand up, Walking</i> }.
Actividad	La actividad que el residente realiza { <i>Anomaly, Check over, Toilet, Inside, Outside, Medical attention, Medication, Hygiene, Mood, Recreation, Social interactions, Feeding, Visits, Vitals</i> }.
Detalles	Descripción detallada sobre la actividad que efectúa el residente.

respectivamente. En la campaña 2 se utilizó el sistema implementado por la residencia basado en los prototipos mencionados anteriormente.

Tabla 4. Características generales del repositorio de datos obtenido en una residencia geriátrica.

	Campaña de sensado 1 (CS1)	Campaña de sensado 2 (CS2)
Periodo de colección de datos	Julio, 2012 a Enero, 2013	Febrero, 2013 a Noviembre, 2015
Número de meses	6	33
Número de participantes	15	40 ³
Número de días	129	954
Número total de registros colectados	19,715	184,806

La campaña 1 se realizó durante seis meses, dentro del periodo de los meses de Julio de 2012 a Enero de 2013, equivalentes a 129 días; con un número de 15 participantes, quienes generaron cerca de 20 mil registros. La colección de datos para la campaña 2 fue durante el periodo de Febrero de 2013 a Noviembre de 2015, es decir, 33 meses; contando con la participación de 40 residentes durante 954 días y obteniendo un total aproximado de 185 mil registros. Los detalles de cada repositorio de datos se describen en la Tabla 4.

Las características demográficas de los participantes residentes en el periodo de colección de datos se muestran en la Tabla 5. Los nombres tanto de residentes como de cuidadores fueron etiquetados con un identificador, como medida de protección

³Únicamente 10 residentes participaron en ambas campañas de colección de datos.

de datos personales. La población de participantes está constituida por el 49% de mujeres y 51% de hombres. La edad promedio es 81.49 años ($ds = 9.65$). Los principales padecimientos entre los participantes son la demencia, Alzheimer, Parkinson y deterioro físico.

Tabla 5. Características demográficas y de salud de los participantes.

Residente	Edad	Género	Padecimientos ⁴	Residente	Edad	Género	Padecimientos
01	80	Hombre	ML, PrCa, PrAu	24	85	Mujer	D
02	90	Mujer	A	25	65	Mujer	-
03	92	Mujer	A	26	56	Hombre	-
04	89	Hombre	A, ML, Db	27	70	Mujer	D
05	90	Hombre	P	28	74	Mujer	D
06	89	Mujer	D, DF	29	82	Hombre	D
07	89	Hombre	ML	30	65	Hombre	-
08	71	Hombre	P	31	75	Hombre	D
09	94	Mujer	D, ML	32	78	Hombre	D
10	92	Mujer	ML, PrRe	33	99	Hombre	D
11	78	Mujer	D	34	76	Mujer	D
12	85	Hombre	C, DF	35	91	Mujer	D
13	90	Hombre	ML, PrAu	36	76	Mujer	D
14	80	Hombre	DF	37	81	Mujer	A
15	92	Hombre	A, ML	38	68	Mujer	-
16	88	Mujer	D	39	83	Mujer	D
17	90	Mujer	D	40	82	Hombre	-
18	80	Mujer	A	41	57	Hombre	-
19	85	Mujer	-	42	85	Mujer	-
20	89	Hombre	-	43	84	Hombre	D
21	75	Hombre	A	44	76	Hombre	-
22	90	Mujer	-	45	76	Hombre	-
23	85	Hombre	-				

La distribución del conjunto de datos por cada participante se puede observar en la Tabla 6. De donde el promedio de días por participante fue de 91 y 312 días para la campaña 1 (CS1) y la campaña 2 (CS2) respectivamente. El promedio del número máximo de registros por día fue de 25 en la CS1 contra 33 en la CS2. Por último del promedio de registros por día para la CS1 fue de ocho y para la CS2 fue de 12.

⁴D = Demencia, A = Alzheimer, P = Parkinson, ML = Movilidad limitada, Db = Diabetes, PrCa = Problemas cardiacos, PrRe = Problemas respiratorios, PrAu = Problemas auditivos, DF = Deterioro físico.

Tabla 6. Características del repositorio por cada participante (CS1 = campaña de sensado 1 y CS2 = campaña de sensado 2 respectivamente).

Participante	Núm. total de días		Núm. total de reg.		Prom. de reg. por día		Núm. máx. de reg. por día	
	CS1	CS2	CS1	CS2	CS1	CS2	CS1	CS2
	CS1	CS2	CS1	CS2	CS1	CS2	CS1	CS2
01	125	905	1,919	13,282	15.35	14.68	55	44
02	123	-	1,283	-	10.43	-	28	-
03	118	909	1,154	14,122	9.78	15.54	36	48
04	126	671	2,421	9,350	19.21	13.93	55	49
05	125	7	2,066	8	16.53	1.14	51	2
06	126	886	1,935	13,693	15.36	15.45	53	56
07	113	860	531	9,178	4.70	10.67	16	42
08	94	902	2,589	16,733	27.54	18.55	58	55
09	126	910	1,503	13,258	11.93	14.57	35	45
10	122	119	1,529	615	12.53	5.17	42	30
11	66	-	974	-	14.76	-	33	-
12	24	-	351	-	14.63	-	31	-
13	51	-	339	-	6.65	-	14	-
14	31	-	587	-	18.94	-	33	-
15	1	278	1	2,261	1	8.13	1	28
16	-	24	-	162	-	6.75	-	23
17	-	789	-	12,785	-	16.20	-	39
18	-	883	-	13,375	-	15.15	-	51
19	-	86	-	142	-	1.65	-	4
20	-	585	-	2,495	-	4.26	-	17
21	-	567	-	5,584	-	9.85	-	33
22	-	15	-	231	-	15.40	-	30
23	-	6	-	108	-	18.00	-	40
24	-	212	-	4,267	-	20.13	-	41
25	-	7	-	70	-	10.00	-	22
26	-	296	-	4,468	-	15.09	-	41
27	-	53	-	849	-	16.02	-	37
28	-	394	-	7,233	-	18.36	-	51
29	-	26	-	295	-	11.35	-	33
30	-	27	-	471	-	17.44	-	31
31	-	438	-	8,067	-	18.42	-	46
32	-	13	-	66	-	5.08	-	7
33	-	3	-	14	-	4.67	-	8
34	-	111	-	2,730	-	24.60	-	58
35	-	18	-	364	-	20.22	-	33
36	-	509	-	9,063	-	17.81	-	48
37	-	10	-	102	-	10.20	-	19
38	-	475	-	8,017	-	16.88	-	50
39	-	3	-	18	-	6.00	-	9
40	-	37	-	509	-	13.76	-	42
41	-	4	-	4	-	1.00	-	1
42	-	15	-	228	-	15.20	-	28
43	-	237	-	4,957	-	20.92	-	47
44	-	78	-	1,509	-	19.35	-	36
45	-	130	-	1,342	-	10.32	-	28

3.5. Resumen

En este capítulo se presentó el repositorio de datos empleado para el desarrollo de la tesis. Se proporcionaron los detalles de los tipos de datos y sus características,

así como de los participantes. A partir de estos datos se obtuvo una representación denominada registro de actividad, en donde se engloban todos los tipos de datos contenidos en el repositorio. En el siguiente capítulo se presentará la forma en la cual estos datos son analizados para el descubrimiento de patrones y anomalías a partir de las actividades del adulto mayor.

Capítulo 4. Componente híbrido para el análisis del registro de actividad

En este capítulo se presenta un componente híbrido para el análisis del registro de actividad propuesto en el capítulo anterior. Este componente está basado en dos perspectivas para análisis de los diferentes tipos de datos contenidos en el repositorio: la perspectiva guiada por datos o *data-driven approach (DDA)*, y la perspectiva guiada por conocimiento o *knowledge-driven approach (KDA)*, ambas expuestas en el capítulo 2.

4.1. Descripción general

El componente híbrido para el análisis del registro de actividad (Soto-Mendoza *et al.*, 2017) propuesto en este trabajo de tesis se presenta en la Figura 4. Al inicio del proceso, como entrada, se tiene el repositorio de origen naturalista conteniendo los diferentes tipos de datos presentados en el capítulo 3. Todos y cada uno de los datos pasan por un pre-procesado (*pre-processing*) antes de ser analizados. Después del pre-procesado se genera un conjunto de registros de actividad. Con los registros de actividad se alimentan los diferentes repositorios de datos: *training data set (TDS)*, *queried data set (QDS)* y el *TBox* parte de la base de conocimientos o *knowledge base (KB)*. El *TBox* es el conjunto de registros de actividad materializados en instancias OWL (Ontology Web Language, por sus siglas en inglés).

Los módulos de análisis y predicción utilizan las perspectivas *DDA* y *KDA* para realizar sus procedimientos correspondientes. En el módulo de análisis se desarrolla de manera *a priori* un análisis sobre el registro de actividad empleando técnicas de aprendizaje automático y técnicas semánticas. Los resultados del módulo de análisis pueden pasar ya sea al módulo de visualización para su presentación gráfica o al módulo de predicción para seguir siendo procesados antes de ser visualizados.

Finalmente, el módulo de visualización muestra los datos a través del uso de técnicas para la visualización de datos. Cabe mencionar que los procesos de análisis, predicción y visualización pueden realizarse considerando diferentes periodos de tiempo (por ejemplo, por registro, por día, por semanas, por meses, etcétera). Finalmente, el

usuario, en este caso los cuidadores, médicos, enfermeras y familiares del adulto mayor pueden observar los cambios en comportamiento a través del tiempo, al mismo tiempo que se les notifica de cualquier desviación en el comportamiento que se detecte. A continuación se explicará a detalle cada uno de los elementos que conforman el componente híbrido propuesto para el análisis del registro de actividad.

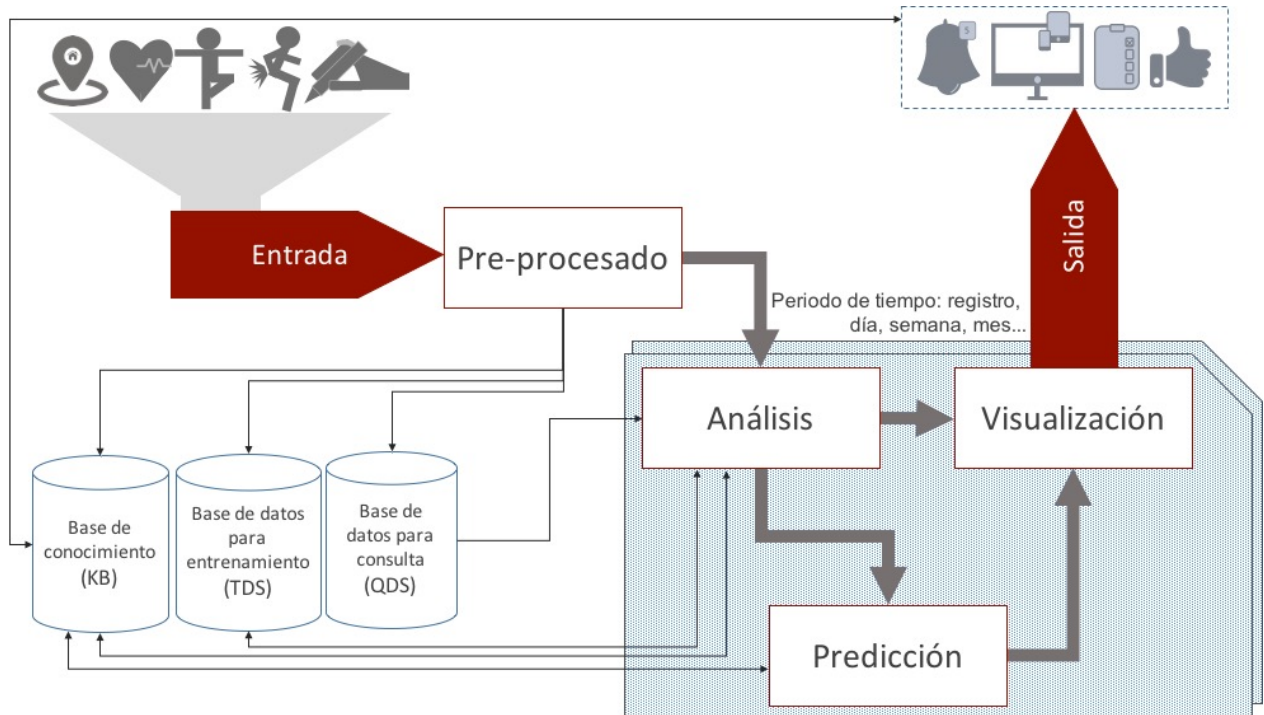


Figura 4. Diagrama del componente híbrido para el análisis del registro de actividad.

4.2. Pre-procesado

El pre-procesado de los datos es un proceso iterativo y tedioso que involucra la participación de un experto o curador de datos. Este proceso involucra una limpieza del repositorio de datos naturalista. Es decir, hasta este punto los datos quedaron estructurados y estandarizados para el análisis en las siguientes etapas. El pre-procesado realizado incluye las siguientes tareas:

- Limpieza. Los errores gramaticales y tipográficos fueron eliminados de los datos. La Tabla 7(a) muestra un ejemplo de los datos originales (en la columna izquierda) donde se puede apreciar que los datos se registraron en el idioma español separando cada palabra mediante signos de puntuación o incluso algunos enun-

ciados y palabras están incompletos. En la columna derecha de la tabla se exhibe la corrección realizada sobre el registro de ejemplo mostrado.

- Eliminación de duplicados. Dentro del repositorio se encontraron registros que contenían la misma información, los cuales fueron eliminados. Comúnmente estos registros aparecen de manera consecutiva en el repositorio, por lo que una forma de detectarlos es a través de la cercanía entre sus respectivas marcas de tiempo. Otra manera para detectar duplicados es comparando las actividades. En el ejemplo mostrado en la Tabla 7(b) la actividad de *Alimentar* está incompleta y se generó un nuevo registro para completar la información.
- Expansión de registros. En algunas ocasiones se incluyeron anotaciones libres de los cuidadores, es decir sin estructura, por lo que un solo registro puede contener más de una actividad. En estos casos, el registro se separó en varios registros de tal forma que una sola actividad quedara indicada en cada registro. En el ejemplo mostrado en la Tabla 7(c), la residente *Jhoan* recibió visitas de sus familiares; lo que implica la propia visita pero también que la residente tuvo interacciones sociales con sus familiares. Por lo tanto fueron creados dos registros: uno para las visitas (*Visits*) y otro para la interacción social (*Social Interactions*).
- Anonimización. La privacidad es fundamental para hacer disponible el repositorio de datos, todos los participantes estuvieron de acuerdo en participar en la colección de datos y firmaron un documento de confidencialidad. Por esta razón los nombres tanto del personal de la residencia como de los residentes, médicos y familiares fueron modificados (ver Tabla 7(d)).
- Traducción. Como se ha visto, la mayoría de los registros del repositorio de datos fueron originalmente capturados en el idioma español (en algunos casos en inglés). Debido a esto y con la finalidad de homogenizar los datos del repositorio, todos los registros fueron traducidos al idioma inglés (ver Tabla 7(e)).
- Materialización. Para realizar el análisis semántico los datos del repositorio fueron semantificados a instancias OWL usando OWL-API en RDF/XML (ver Tabla 7(f)). Estos registros materializados son el conjunto de instancias que alimentan a la base de conocimientos (KB), específicamente es lo que se conoce como *TBox*.

Los datos pre-procesados se dividieron en dos conjuntos de datos: el de entrena-

Tabla 7. Etapas del pre-procesado.

Datos originales	Datos pre-procesados
a) Limpieza	
5/8/2013, 15:36:52, 4, Alejandro, Elena, Estado de animo, Tranquilo: Se. En Cuenta. De. Buen. Animo. Accecible. Al. Manejo.,cesi. Le. Dio. Maje. Por. Una. Hora. Lo. Cual. Copera. Muy. Bien.	5/8/2013, 15:36:52, 4, Caregiver06, Resident03, Mood, Cheerfull
b) Eliminación de duplicados	
5/24/2013, 15:39:45 , 6, Susan, Lily, Alimentar, Desayuna Y Come Muy Bien 5/24/2013, 15:39:46 , 6, Susan, Lily, Alimentar	5/24/2013, 15:39:45,6, Caregiver03, Resident26, Feeding, Breakfast:100
c) Expansión de registros	
10/27/2015, 6:07:00pm, 3, Jhoan, Maria, Visitas, tiene,la visita de su hijas sus nietos y su padre	10/27/2015, 18:07:00, 3, Caregiver08, Resident26, SocialInteractions, Family 10/27/2015, 18:07:00, 3, Caregiver08, Resident26, Visits, Children, Grandchildren, Parents
d) Anonimización	
8/5/2013, 16:06:10, 2, Resident06, Ana , Visitas, 1:50pm Recibe visita de su hija Amanda regresando a casa felicidad a las 3pm 8/6/2013, 16:33:55, 3, Resident06, Lily, Salidas, Salida a cita medica con doctor Zamudio	8/5/2013, 13:50:00, 2, Caregiver09, Resident06, SocialInteractions, Family 8/5/2013, 13:50:00, 2, Caregiver09, Resident06, Visits, Children 8/5/2013, 15:00:00, 2, Caregiver09, Resident06, Inside, Back 8/6/2013, 16:33:55, 3, Caregiver03, Resident06, Outside, Medical
e) Traducción	
5/10/2013, 15:42:58, 6, Resident03, Charity, Aseo Personal, Showered 5/26/2013, 17:22:20, 1, Resident03, Gloria, Aseo Personal, Toma baño en regadera y arreglo personal	5/10/2013, 15:42:58, 6, Caregiver05, Resident03, Hygiene, Shower 5/26/2013, 17:22:20, 1, Caregiver03, Resident03, Hygiene, Shower
f) Materialización	
569, 7/21/2012, 7:12:19, 7, Caregiver04, Resident04, Mood, Annoyed	:e569 a ontBeh:Record . :e569 ontBeh:hasDate "2012-07-21T09:12:19"^^ xsd:dateTime . :e569 ontBeh:createBy Caregiver04 . :e569 ontBeh:concernsResident :Resident04 . :e569 ontBeh:hasMood annoyed .

miento (*training dataset, TDS*) y el de consulta (*query dataset, QDS*). El *TDS* se utiliza para entrenar los métodos y técnicas descritas en las siguientes secciones. El *QDS* se utiliza para probar y hacer la experimentación con el componente híbrido propuesto, eso se presentará en el Capítulo 5. Teniendo en cuenta las diferencias en el número de días de cada participante entre cada campaña de sensado, el conjunto de datos se dividió de acuerdo con la teoría de aprendizaje de Abu-Mostafa *et al.* (2012), donde el

conjunto de datos de prueba es $K = \frac{N}{5}$ y N es el tamaño total del conjunto de datos. En este caso cada residente tiene un N diferente como se puede ver en la Tabla 8.

Tabla 8. Número de días y registros utilizados como datos de entrenamiento y prueba por residente.

Residente	Número de días		Número de registros		Residente	Número de días		Número de registros	
	Entrena- miento	Prueba	Entrena- miento	Prueba		Entrena- miento	Prueba	Entrena- miento	Prueba
01	822	206	10802	4358	19	69	17	110	32
02	98	25	1034	249	20	469	117	2102	416
03	821	205	10530	4770	21	454	113	4491	1102
04	636	159	9576	2369	24	170	42	3197	991
05	106	26	1827	247	26	238	59	3481	971
06	810	202	11137	4568	27	42	11	694	152
07	778	195	6464	3231	28	315	79	5235	1945
08	797	199	14098	5322	31	350	88	5711	1845
09	829	207	10228	4598	34	89	22	2093	629
10	193	48	1743	403	36	407	102	6483	2465
11	53	13	830	144	38	380	95	5743	2231
13	41	10	288	51	43	190	47	3877	1045
15	223	56	1622	663	44	63	16	1236	266
17	632	158	9162	3543	45	104	26	1016	316
18	707	177	9090	4225					

4.3. Análisis

En la etapa anterior todo el conjunto de datos fue homogenizado a una sola estructura. Después de finalizar la tarea de pre-procesado se redujo el número de registros obtenidos originalmente. Las características del repositorio de datos pre-procesados se muestra en la Tabla 9. En la tabla se observa claramente que el número total de registros en la *CS2* es mayor comparado con los de la *CS1* debido a la duración del periodo de colección de datos. Al mismo tiempo el número de registros por día es mayor porque los cuidadores ya estaban familiarizados con el registro constante en el sistema de registro de actividades del adulto mayor.

Tabla 9. Características del repositorio de datos pre-procesado.

	Campaña de sensado 1, <i>CS1</i>	Campaña de sensado 2, <i>CS2</i>
Número total de registros	19,182	163,157
Promedio de número de días por participante	91	312.50
Promedio de número máximo de registros por día	25	33.80
Promedio de número de registros por día	8.63	12.95

En esta etapa, los datos ya están listos para ser analizados desde dos perspectivas diferentes: la guiada por datos (*DDA*) y la guiada por conocimiento (*KDA*). Ambas perspectivas combinadas son incluidas en diferentes fases del análisis complementándose entre sí.

4.3.1. Análisis basado en datos

En esta parte del análisis se utilizaron técnicas de minería de datos y aprendizaje automático para obtener información estadística de cada adulto mayor. Las frecuencias y promedios obtenidos a partir del *TDS* permitieron establecer una línea base del comportamiento normal en cada actividad para cada uno de los residentes. El resumen de las estadísticas obtenidas se muestran en la Tabla 10.

Tabla 10. Estadísticas obtenidas a partir del *TDS*.

Hora de dormir	Número de comidas al día	Rango de signos vitales
Horas de sueño	Número de visitas	Número de visitas nocturnas al baño
Hora de ocurrencia por actividad	Secuencia de actividades frecuentes	

Clúster. Este método se utilizó para efectuar un análisis basado en tiempo. El algoritmo trata de crear grupos a partir de un conjunto de datos y es utilizado para la detección de anomalías. De manera general el funcionamiento del algoritmo es el siguiente: se define el número de grupos que se desea crear. Cada grupo está representado por un centroide del cual se determina su valor. Posteriormente se calculan las distancias de todos los puntos a cada centroide, y se asigna cada punto al centroide más cercano. Se ajustan los valores del centroide y se vuelve a repetir el proceso, hasta que se cumpla la condición de paro.

Trasladando el problema a los datos del adulto mayor, se consideró el conjunto de actividades por residente y su hora de ocurrencia para determinar los grupos aplicando la técnica de clúster. Así se identificaron los horarios habituales de cada actividad por cada residente, es decir, se obtuvieron las horas del día más frecuentes donde cada residente realizaba cada una de sus actividades. El algoritmo de k-medias (Hartigan y Wong, 1979) detecta el centroide para cada actividad (ver Figura 5). Algunas de ellas se ejecutan varias veces al día por lo que se obtuvo un conjunto de centroides. Para los detalles del funcionamiento de esta técnica revisar el Anexo B.

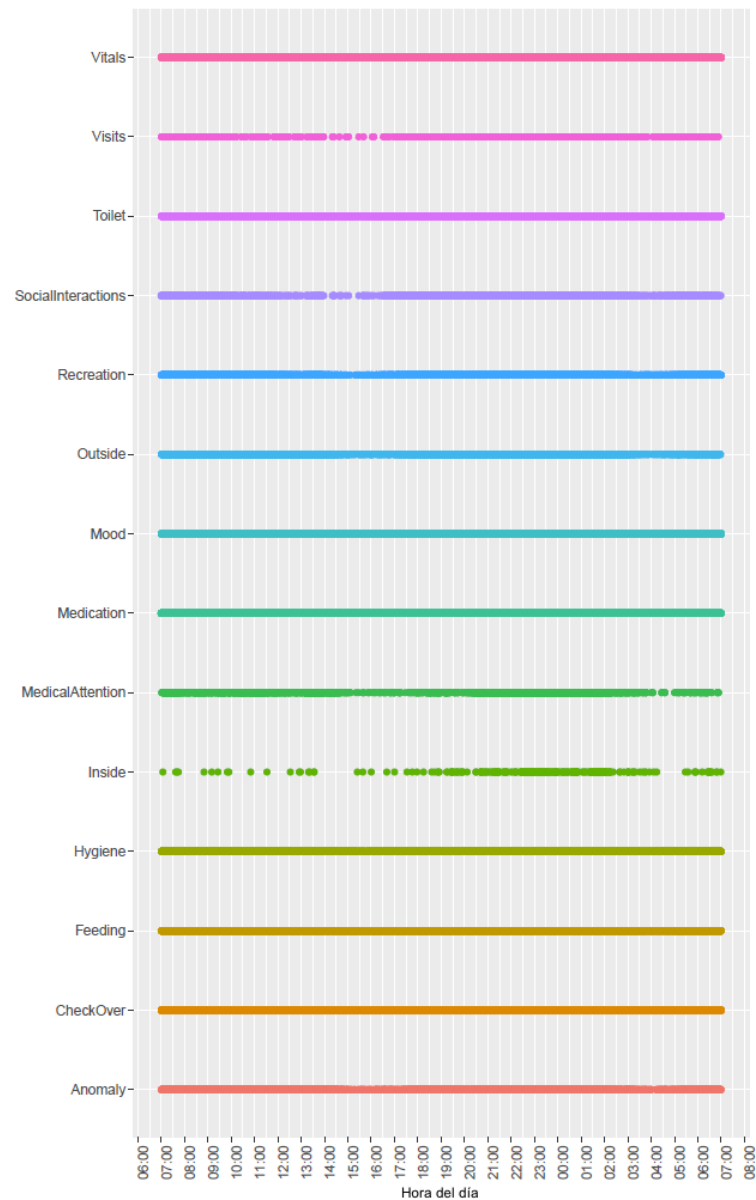


Figura 5. Análisis del tiempo de ocurrencia de cada actividad.

Minería de patrones secuenciales. Esta técnica realiza un análisis de patrones a nivel de secuencia de actividades (Chand *et al.*, 2012), complementando el análisis basado en tiempo. Donde dada una lista de secuencias compuestas por transacciones ordenadas cronológicamente, el objetivo es descubrir todos los patrones secuenciales que aparecen en un cierto número de secuencias.

Trasladando este problema a los datos que conforman el repositorio, se utiliza como entrada la secuencia diaria de actividades con una ligera modificación: las actividades se intercambiaron por símbolos respetando el orden cronológico. Las secuencias son

las secuencias diarias de actividades y las transacciones son las actividades. Se define una longitud k de secuencia, se buscan secuencias de tamaño k , y se contabiliza el número de secuencias en el que aparece ese patrón. Como resultado se tiene una lista de secuencias más frecuentes con distintas longitudes. De esta manera, se calcularon las secuencias de actividades más frecuentes para cada residente dentro de un periodo de tiempo (ver Figura 6). Para los detalles del funcionamiento de esta técnica revisar el Anexo B.

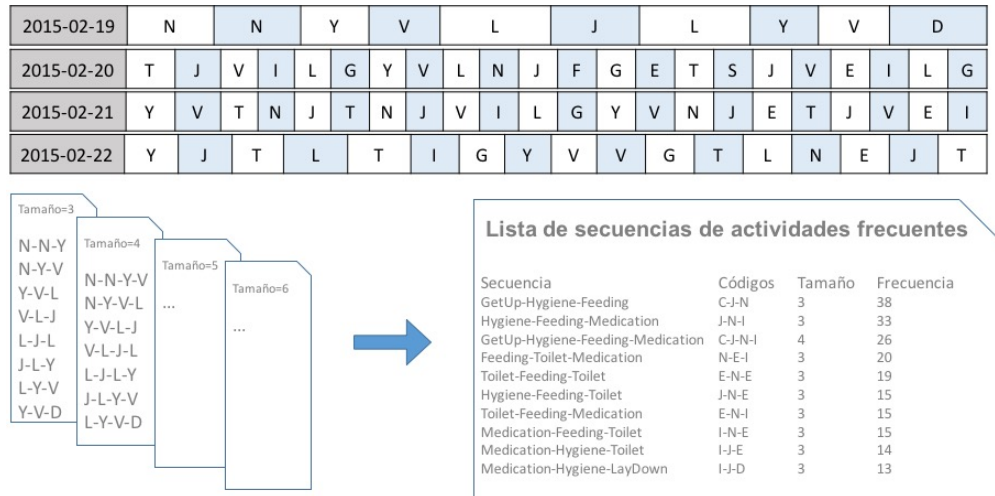


Figura 6. Análisis de las secuencias de actividades frecuentes.

Compresión de la secuencia de actividades. Para este análisis se seleccionó un método de compresión basado en gramáticas. *Re-pair* es un esquema de compresión de datos que es altamente eficiente, particularmente en descompresión (Larsson, 1999).

Las actividades pueden ser vistas como una secuencia de símbolos. La secuencia corresponde a un día completo de un residente, en donde las actividades están ordenadas cronológicamente. Cada símbolo representa un átomo (una actividad, en un lugar, en un momento dado, etc.). Se asigna un símbolo a cada átomo y se cuentan *pares* de símbolos. El par más frecuente es sustituido por un nuevo símbolo, y el proceso se repite hasta que sólo queda un símbolo (en el mejor caso). En cada paso, se produce al menos una fusión. El procedimiento anterior produce un árbol, o en este caso, una pirámide de símbolos (ver Figura 7). El objetivo tras aplicar esta técnica es reducir la secuencia de actividades, acelerar el análisis y la visualización (como se ve-

rá en la sección 4.5). Para los detalles del funcionamiento de esta técnica revisar el Anexo B.

Iteración	Par	Secuencia	Frecuencia máxima	Tamaño de la secuencia
0		AFBCDBCDFGJIAFCDLBDFLG	CD = 3	22
1	a → CD	AFBaBaFGJIAFaLBDFLG	Ba = 2, AF = 2	19
2	b → Ba	AFbbFGJIAFaLBDFLG	AF = 2	17
3	c → AF	cbbFGJlcaLBDFLG		15

Figura 7. Compresión de la secuencia de actividades.

Las estadísticas obtenidas previamente se obtuvieron a partir del *TDS* y sirvieron para incluir propiedades y tipos de datos a la ontología. La ontología también es parte de la base de conocimientos. Estos datos son relevantes ya que permiten hacer inferencias de comportamiento utilizando el modelo semántico.

4.3.2. Análisis basado en conocimiento

Se desarrolló una base de conocimiento (*KB*) para realizar un análisis semántico de las actividades del adulto mayor. La *KB* está conformada por dos elementos principales: un *TBox*, es decir, una ontología; y por un *ABox*, representado por las instancias de clases y un conjunto de reglas de inferencia. La ontología se define como una especificación formal y explícita de una conceptualización compartida (Studer *et al.*, 1998); la ontología es una herramienta semántica para una representación formal de un contexto específico.

La metodología para el diseño de la ontología sigue una perspectiva de arriba hacia abajo. Es decir, expertos y datos obtenidos del dominio de aplicación participaron en el diseño (Soto-Mendoza y Garcia-Macias, 2014). Los conceptos de la ontología surgieron a partir de preguntas de competencia (Staab *et al.*, 2001). El diseño es modular ya que se incorporaron diferentes vocabularios y otras ontologías disponibles en la literatura. La utilización de vocabularios estándar incrementó la compatibilidad y la reutilización en futuras aplicaciones. Para formalizar la ontología se empleó el lenguaje OWL2 a

través del uso de un perfil de lógica descriptiva y una sintaxis con RDF/XML. Protégé¹ fue el editor de ontologías empleado para este propósito.

Los diferentes estados (*i.e.* *LayDown*) al mismo tiempo que los comportamientos (*i.e.* *mood_behavior*) del adulto mayor fueron modelados en la ontología a través de una jerarquía dividida en dos vocabularios independientes. En este sentido, el vocabulario tiene la facilidad de extenderse. De esta manera, se generó una representación semántica del registro de actividad mostrada en la Figura 8.

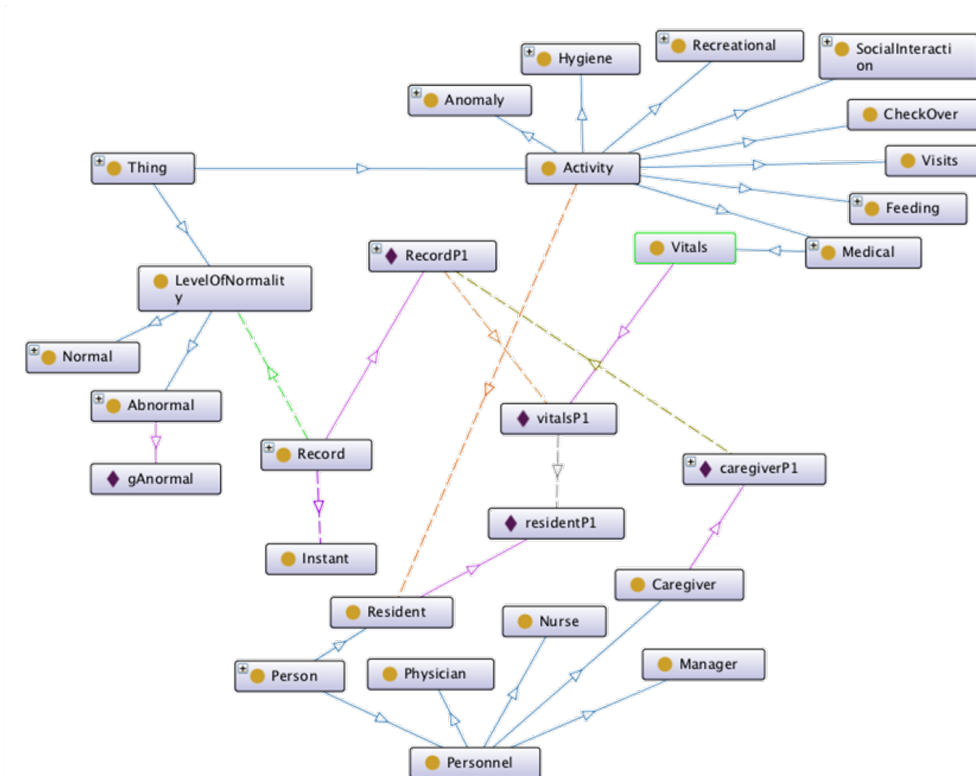


Figura 8. Representación semántica del registro de actividad.

Para la interoperabilidad con otros recursos semánticos disponibles en la web, la ontología FOAF², una de las más populares, fue extendida para describir a los residentes y al personal de la residencia geriátrica. También la ontología OWLtime³ fue reutilizada para modelar las propiedades temporales del registro de actividad. Además, algunos conceptos de la ontología de alto nivel CONON (Wang *et al.*, 2004) fueron incluidos. El diagrama general de la ontología denominada *ABIOAOntology* se presenta en la Figura 9.

¹<http://protege.stanford.edu>

²<http://xmlns.com/foaf/spec/>

³<https://www.w3.org/TR/owl-time/>

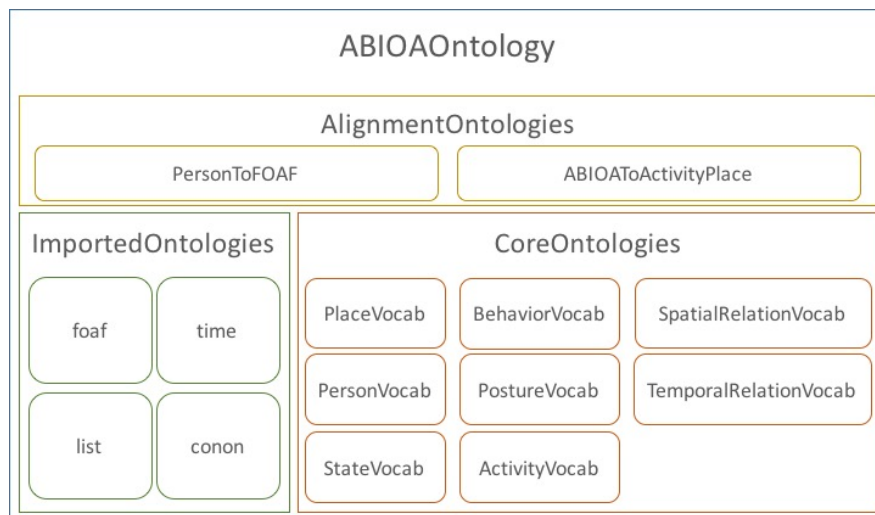


Figura 9. Diagrama general de *ABIOAOntology*.

4.4. Predicción

En este componente se realiza una detección de anomalías desde varias perspectivas: una basada en tiempo, otra basada en secuencia y la última mediante un proceso de razonamiento semántico usando *ABIOAOntology*. Para esto se definieron umbrales para los niveles de anomalía denominados como *normal*, *ligeramente anormal*, *medianamente anormal* y *severamente anormal* dependiendo del tipo de predicción que se aplique. Con estos niveles de anomalía y definiendo umbrales, es posible detectar actividades y días anormales en las rutinas de los adultos mayores. Los umbrales fueron determinados de manera empírica a partir de observaciones y entrevistas con los dueños del proceso. De forma análoga, en la literatura existen precedentes de la definición de umbrales empíricos para medir anomalías (Mahmoud *et al.*, 2011; Candás *et al.*, 2014; Moshtaghi *et al.*, 2015).

4.4.1. Detección de anomalías basadas en tiempo.

Con los centroides obtenidos es posible determinar de manera personalizada primero si una actividad es normal o anormal. Después procesando todas las actividades de un día, si un día es normal o anormal, con base en el tiempo de ocurrencia de la actividad. Los niveles de anomalía y sus respectivos umbrales se muestran en la Tabla 11. En este caso, el umbral representa la distancia a partir del centroide. Por ejemplo, si la hora del desayuno (*Feeding*) del *Residente 05* es frecuentemente a las 8:30 ho-

ras, y hoy desayuna a las 9:00 horas, la actividad será clasificada como ligeramente anormal.

Tabla 11. Niveles de anormalidad definidos para el análisis basado en tiempo.

Nivel		Rango
Normal (N)		$x \leq 15$ minutos
Ligero (L)		$15 \text{ minutos} < x \leq 30$ minutos
Anormal	Medio (M)	$30 \text{ minutos} < x \leq 60$ minutos
	Severo (S)	$x > 60$ minutos

4.4.2. Detección de anomalías basadas en secuencia.

La detección de anomalías basada en tiempo se complementa con la detección basada en secuencia, en la cual las secuencias de actividades más frecuentes representan los comportamientos habituales del residente. La lista que contiene las secuencias de actividades más frecuentes se ordena y se almacena. Dentro de esta lista, las secuencias con mayor frecuencia fueron etiquetadas como *Normal* y aquellas con frecuencias menores como *Severamente anormal*.

Para el caso de las actividades del adulto mayor, se necesita comparar una nueva secuencia de actividades con el listado de secuencias de actividades más frecuentes. Es decir, cada nueva secuencia que se desea clasificar se compara con las secuencias de la lista utilizando el algoritmo de la subsecuencia común más larga (LCS, por sus siglas en inglés longest common subsequence) (Bergroth *et al.*, 2000a). Una forma de resolver este problema es empleando la programación dinámica, en donde se van comparando uno a uno los símbolos que conforman las secuencias de actividades a clasificar. Para realizar la clasificación de las secuencias de actividades conforme los diferentes niveles de anomalía se definieron también umbrales (ver Tabla 12). En este caso, el umbral está definido con respecto al porcentaje de similaridad entre la secuencia a evaluar y la secuencia más frecuente. Para los detalles del funcionamiento de esta técnica revisar el Anexo B.

Los niveles de anomalía basados en el tiempo y en secuencia se pueden extraer considerando diferentes ventanas de tiempo. Por ejemplo, es común que un adulto mayor de repente cambie su comportamiento, y algo que fue etiquetado como *Normal* después de dos meses ya no es así. Por lo tanto, el nivel *Normal* debe ser restablecido

Tabla 12. Niveles de anormalidad definidos para el análisis basado en secuencia.

Nivel		Similaridad
Normal (N)		$x \geq 75\%$
	Ligero (L)	$50\% \leq x < 75\%$
Anormal	Medio (M)	$25\% \leq x < 50\%$
	Severo (S)	$x < 25\%$

con la nueva información.

4.4.3. Detección de anomalías con inferencias semánticas

Las detecciones anteriores indican si una actividad o un día es normal o anormal. Sin embargo, hay conocimiento que no se consideró en las inferencias anteriores. Por ejemplo, para los cuidadores y geriatras los signos vitales son un indicador muy importante y cualquier medición fuera de rango ya es signo de que algo anormal ocurre (Chester y Rudolph, 2011). Este hecho impacta directamente el resultado de la inferencia de anomalía realizada, donde un día que fue detectado como normal debe ser anormal porque la presión arterial está fuera de rango, o viceversa. De manera general, las inferencias presentadas en este trabajo pueden enriquecerse y ser más precisas si se considerara información contextual adicional.

4.5. Visualización

La visualización de los datos y de las inferencias realizadas por las técnicas de análisis presentadas previamente es muy importante para los cuidadores, enfermeros, médicos y familiares, y para el mismo adulto mayor; pero también resulta interesante para los analistas de datos. Para estos últimos, la visualización permite identificar tendencias y patrones de forma rápida y sencilla. Por otro lado, debido al volumen de los datos se hace necesaria una representación gráfica abstracta y suficientemente enriquecida con información para la toma de decisiones por parte del personal encargado del bienestar del adulto mayor (Le, 2014). Es importante contar con una representación visual de los datos cuando se busca que éstos repercutan en el comportamiento de los usuarios a largo plazo (Chiang *et al.*, 2014). La visualización de información médica es un reto porque incorpora datos provenientes de distintas fuentes (Shneider-

man *et al.*, 2013). Adicionalmente, el tiempo es una variable crítica cuando se trata del bienestar de una persona. Más aún, desde la perspectiva del negocio, es decir, para los administradores y tomadores de decisiones de la residencia geriátrica, las herramientas de visualización de datos permiten hacer más eficiente el proceso de toma de decisiones (Chen *et al.*, 2012). Debido a todo lo anterior, en este trabajo se diseñaron diferentes formas de presentar y visualizar datos.

4.5.1. Pirámide de actividades

Esta herramienta de visualización muestra en forma de pirámide la secuencia de actividades de un día específico. Los resultados del agrupamiento que efectúa el algoritmo de *re-pair* son utilizados como entrada para esta herramienta de visualización. La Figura 10 muestra las pirámides correspondientes a algunas instancias del *Residente 01*. Cada pirámide corresponde a un día diferente en donde el usuario puede juzgar, a simple vista, si la pirámide corresponde a un día normal o anormal. Por ejemplo, en la figura es claro que el área de la pirámide con fecha del día *2012-10-09* es más pequeña comparada con las cinco pirámides restantes. Esto es debido a que el residente fue sometido a una intervención quirúrgica en el hospital por lo que abandonó la residencia. Este tipo de procedimientos son acontecimientos anormales para los cuidadores ya que determinan un cambio en la rutina seguida hasta el momento, estos cambios pueden ser temporales o permanentes.

4.5.2. Mapas diarios de colores

El objetivo de esta herramienta es que el usuario con un simple vistazo observe un día de actividades y lo compare con otros para que decida si es un día normal o es un día excepcional. Para este fin, se asignó un color diferente a cada tipo de actividad y cada línea de colores corresponde a una secuencia diaria de actividades. A simple vista se pueden observar las similitudes entre días de un único residente o entre varios residentes. En la Figura 11 se muestra el periodo del *2013-07-05* al *2013-07-16* y en el cuadro a la derecha se muestran las actividades y su respectivo color asignado. El número de actividades por día no es fijo, cada día es diferente y depende de las necesidades de cada residente. Sin embargo, se pueden observar las

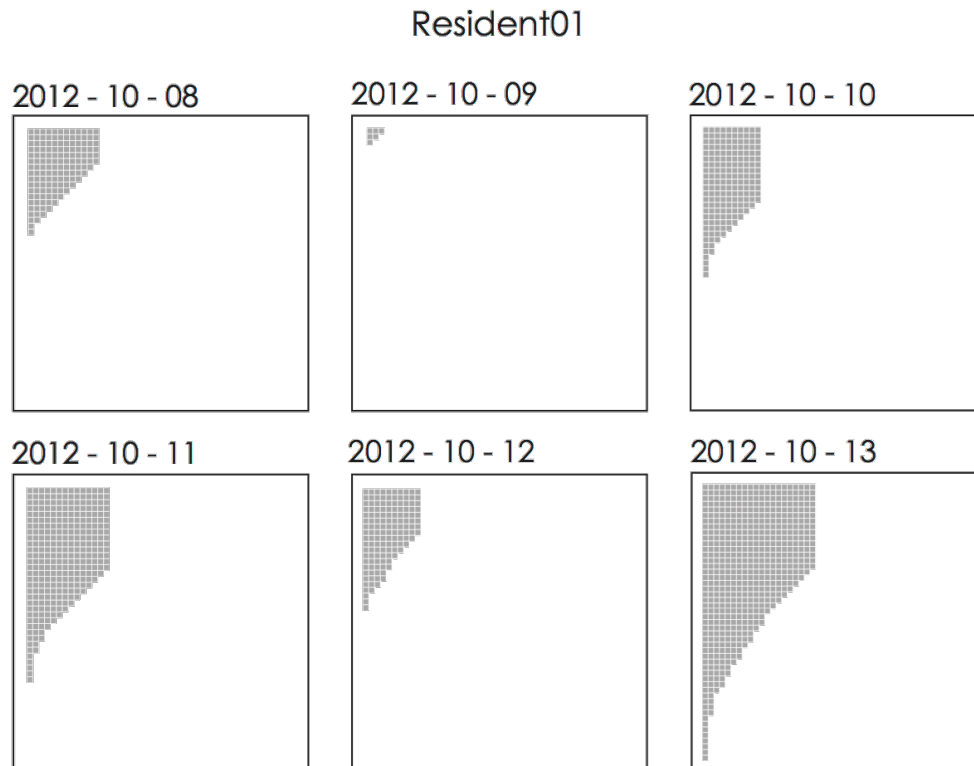


Figura 10. Herramienta de visualización denominada pirámide de actividades.

diferencias entre los días, por ejemplo, los registros de alimentación (*Feeding*) en color gris con los primeros registros en los todos los días. De igual forma, los registros de medicación (*Medication*) aparecen consecutivamente tres veces todos los días, lo que indica que el residente toma tres medicamentos diferentes. Observe que un día los registros de medicación están ausentes, lo que es un indicador de que algo no va como se esperaba. En algunos días los registros de evacuaciones (*Toilet*) no aparecen lo que también es un indicador de que el residente pudiera tener problemas para ir al baño o bien que el cuidador olvidó registrar esta actividad. En cualquier caso, el personal administrativo necesita tomar decisiones al respecto.



Figura 11. Herramienta de visualización denominada mapas diarios de colores.

4.5.3. Línea de actividades

Otra forma de visualizar la secuencia de actividades, es aprovechar la suposición de que las actividades diarias se realizan con cierta periodicidad. Después de todo, los días del adulto mayor guardan cierta rutina (Lymberopoulos *et al.*, 2011); aunque la rutina de cada adulto mayor es diferente. Por lo tanto, las secuencias de actividades frecuentes determinadas anteriormente (ver sección 4.3.1) fueron utilizadas como entrada para esta herramienta de visualización. El nombre del residente aparece del lado izquierdo de cada línea de tiempo. Todas las actividades diarias dentro de un periodo de tiempo de uno o más residentes son mostradas sobre una línea de tiempo (ver Figura 12). Los números sobre las barras de colores indican las fechas (día y mes). El año aparece al posicionar el cursor sobre la fecha. Las secuencias de actividades son mostradas como barras verticales de colores con longitud diferente. El tamaño de la barra representa la longitud de la secuencia de actividades, es decir, el número de actividades consecutivas que forman la secuencia. Se pueden señalar las secuencias de actividades similares a lo largo de toda la línea del tiempo con sólo seleccionar la barra de color. En la figura, aparece un indicador en las barras similares. Asimismo, las actividades que conforman la secuencia de actividades correspondiente a cada barra de color aparecerán al posicionar el cursor sobre la barra. Esta herramienta de visualización permite visualizar datos durante un periodo largo de tiempo. El usuario puede expandir o contraer la línea de tiempo para comparar las secuencias de residentes diferentes y observar los patrones similares. Esta herramienta puede ser utilizada por analistas de datos y por el personal encargado del cuidado del adulto mayor.

4.6. Resumen

Este capítulo presenta el diseño e implementación del componente híbrido para el análisis del comportamiento. Después de que los datos fueron pre-procesados se continuó con el análisis de los mismos siguiendo el paradigma basado en datos y basado en conocimiento. Se obtuvo conocimiento a-priori sobre cada actividad y cada residente que permite establecer una línea base para determinar qué comportamiento es normal y cual no. Se utilizaron las técnicas de clúster, minería de patrones secuenciales, emparejamiento recursivo y subsecuencia común más larga para detectar patrones, todas ellas guiadas por datos. Se diseñó e implementó una ontología, *ABIOAOntology*,

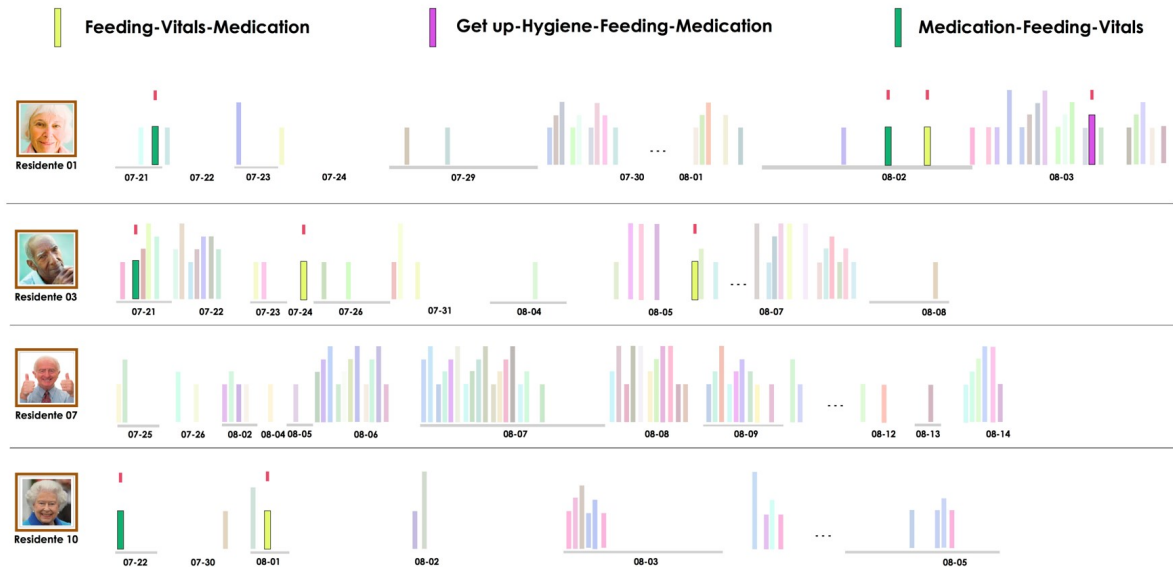


Figura 12. Herramienta de visualización denominada línea de actividades.

para la detección de anomalías con el paradigma basado en conocimiento. Por último, se propone el diseño de tres metáforas para la visualización de los datos: la pirámide de actividades, el mapa diario de colores y la línea de actividades. Los resultados de los experimentos y evaluaciones realizadas con las técnicas y metáforas propuestas, se presentan en el siguiente capítulo.

Capítulo 5. Experimentos y resultados

En este capítulo se presentan los experimentos realizados y se discuten los resultados obtenidos. Se presenta evidencia cuantitativa y cualitativa de la detección automática de anomalías y la evaluación por parte de los usuarios respectivamente.

5.1. Detección automática de anomalías

El objetivo de este experimento es evaluar cada una de las técnicas empleadas para la detección automática de anomalías. Es decir, el análisis basado en tiempo usando clúster, el análisis basado en secuencia empleando la técnica de extracción de patrones secuenciales, por último las inferencias semánticas mediante las reglas de inferencia propuestas. Se seleccionaron los residentes, días y registros del repositorio de datos que cumplían con los siguientes criterios de inclusión: el participante debería tener al menos 30 días de datos de entrenamiento almacenados, el día debe tener más de dos registros de actividades y el número promedio de registros por día debe ser mayor a dos.

Resultados del análisis basado en tiempo. Se utilizaron los datos de entrenamiento para definir los patrones normales de cada residente por cada actividad que realiza, identificando las horas habituales en las cuales se realizaba cada una de éstas. Dado que las horas de un día son períodos cíclicos (Cohen, 2012), se requiere que esta temporalidad quede representada para aplicar la técnica de clúster (ver Anexo B). La representación de las 24 horas del día mostrada en la Figura 13 se logró tras aplicar una transformación a los datos empleando la ecuación 1 (Gröchenig, 2013) sobre los datos de la marca de tiempo correspondiente a cada registro de actividad.

$$\begin{aligned}
 T &= \text{Hora} + (\text{Minuto}/60.0) \\
 X &= \sin\left(\frac{2 * \pi * T}{24}\right) \\
 Y &= \cos\left(\frac{2 * \pi * T}{24}\right)
 \end{aligned} \tag{1}$$

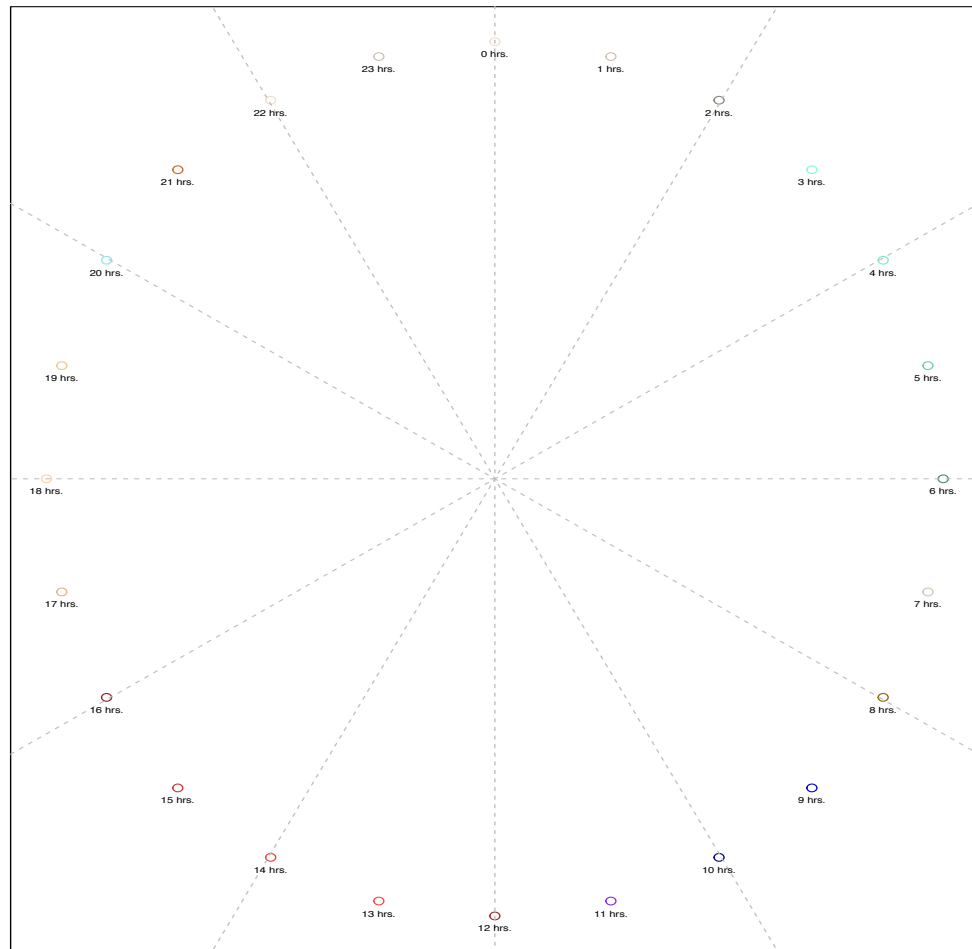


Figura 13. Representación gráfica de las horas a lo largo del día.

En las Figuras 14(a) y 14(b) se puede ver la distribución en el tiempo (hora y minuto) de la actividad de alimentación (*Feeding*) y en las Figuras 15(a) y 15(b) de la actividad de signos vitales (*Vitals*) a lo largo del día, de algunos residentes mostrando únicamente los registros correspondientes a los días de entrenamiento. Por cuestiones de espacio se muestran las más representativas; sin embargo, el análisis se realizó para cada actividad de cada residente.

Posteriormente se organizaron los datos en grupos identificados a través de los centroides después de aplicar la técnica de clúster (Anexo B). En las Figuras 16(a), 16(b), 17(a) y 17(b) se observan los grupos y centroides de las actividades de alimentación y signos vitales para los residentes presentados anteriormente.

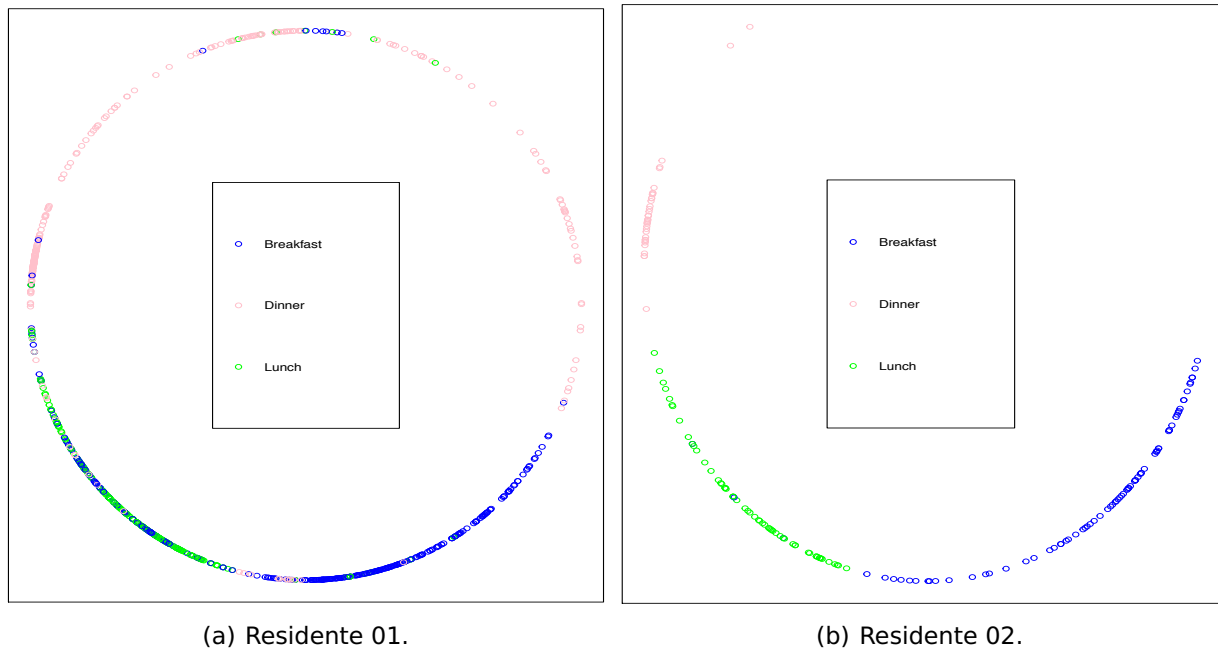


Figura 14. Distribución de las actividades de alimentación (*Feeding*) a lo largo del día.

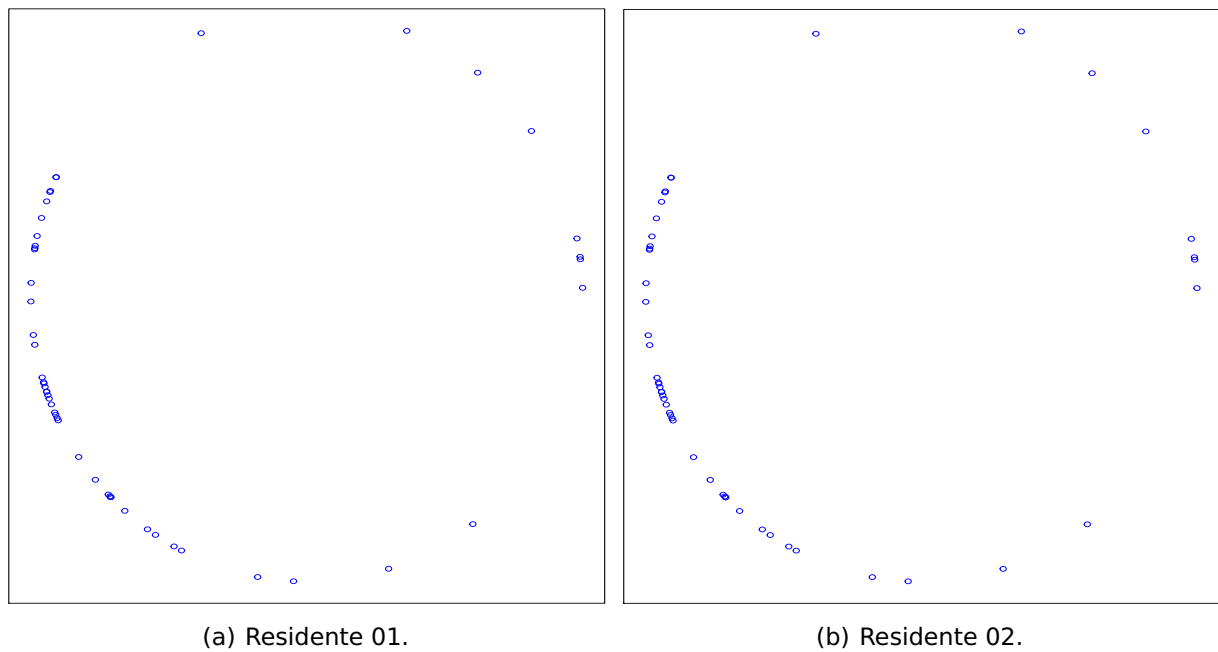


Figura 15. Distribución de las actividades de signos vitales (*Vitals*) a lo largo del día.

Como se puede observar, los centroides de las actividades para cada residente son muy similares. Esto concuerda con los datos obtenidos empíricamente dado que los cuidadores realizan con cada residente la misma actividad casi al mismo tiempo. Por ejemplo, cuando los cuidadores toman los signos vitales, van con cada residente

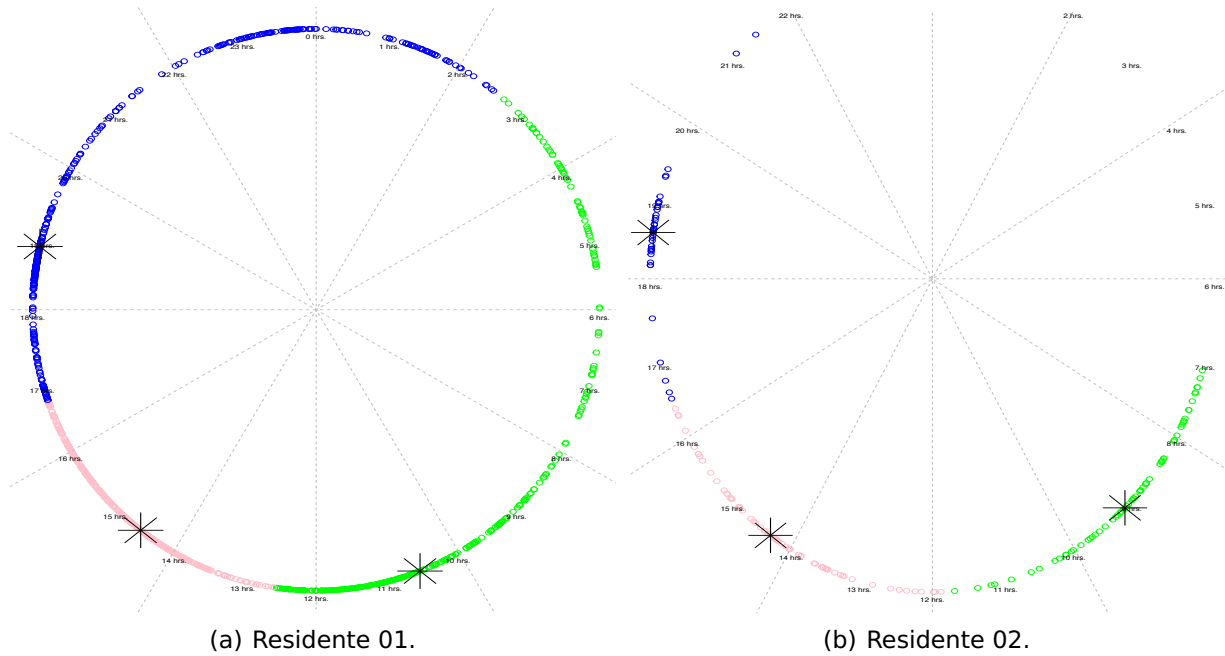


Figura 16. Identificación de horarios habituales de las actividades de alimentación.

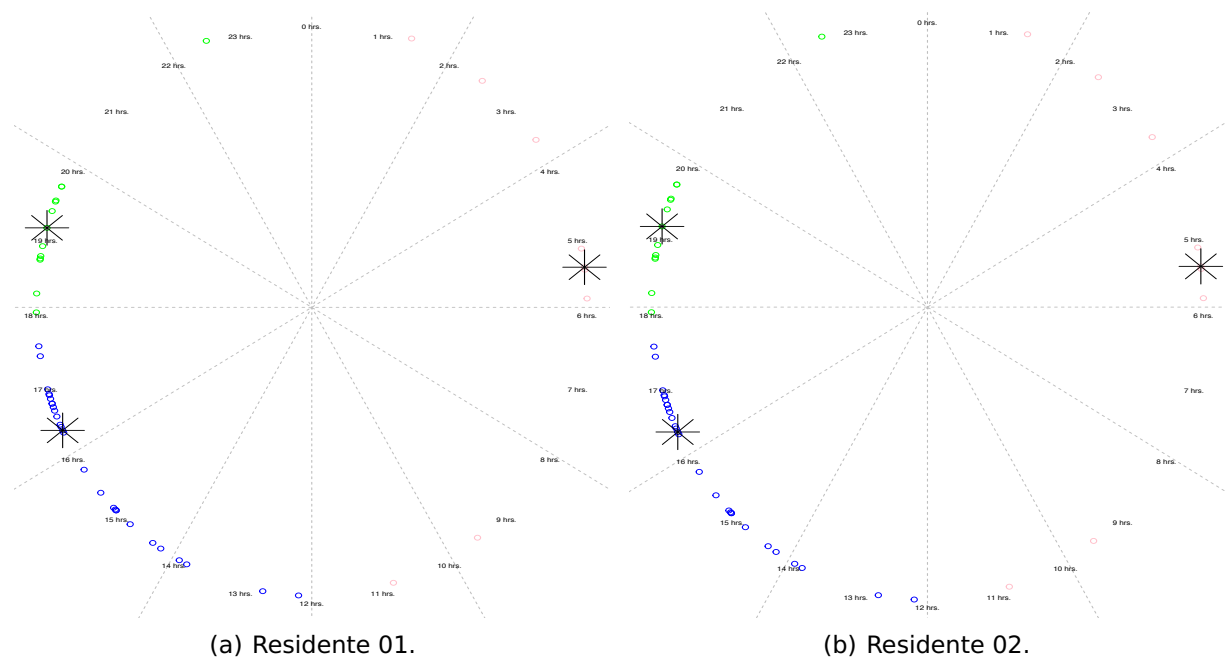


Figura 17. Identificación de horarios habituales de las actividades de signos vitales.

tomando las mediciones correspondientes. Al mismo tiempo, se puede ver que hay algunas diferencias en los horarios en los que los registros quedaron almacenados con respecto a los "horarios oficiales" en los que deberían de aparecer los registros. Esto es, en el caso de los signos vitales, éstos son requeridos de forma obligatoria

tres veces al día para cada residente en horarios perfectamente establecidos (7, 13 y 18 horas); sin embargo, se puede visualizar que los centroides indican otros horarios (5:30, 16:30 y 19:30 horas aproximadamente). Esto pudiera representar una alerta para que los cuidadores tengan mayor cuidado de realizar esta actividad dentro de los horarios exigidos por el médico y el personal administrativo de la residencia.

Una vez que los centroides fueron calculados se obtuvo una clasificación de actividades como normales o anormales después de evaluar el tiempo de ocurrencia de cada actividad de forma personalizada. De igual forma, esto permitió clasificar un día como normal o anormal utilizando los niveles de anormalidad establecidos (ver Tabla 11) midiendo la distancia entre los puntos con respecto a los centroides.

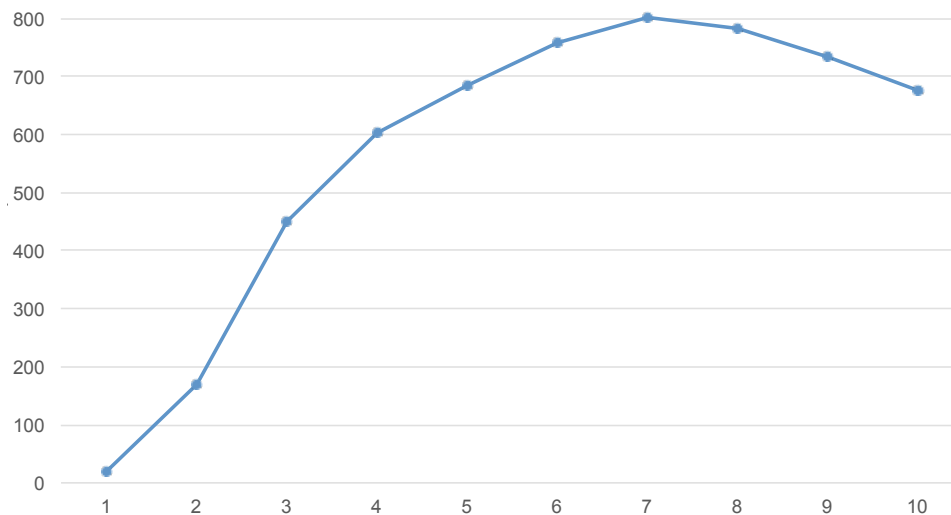
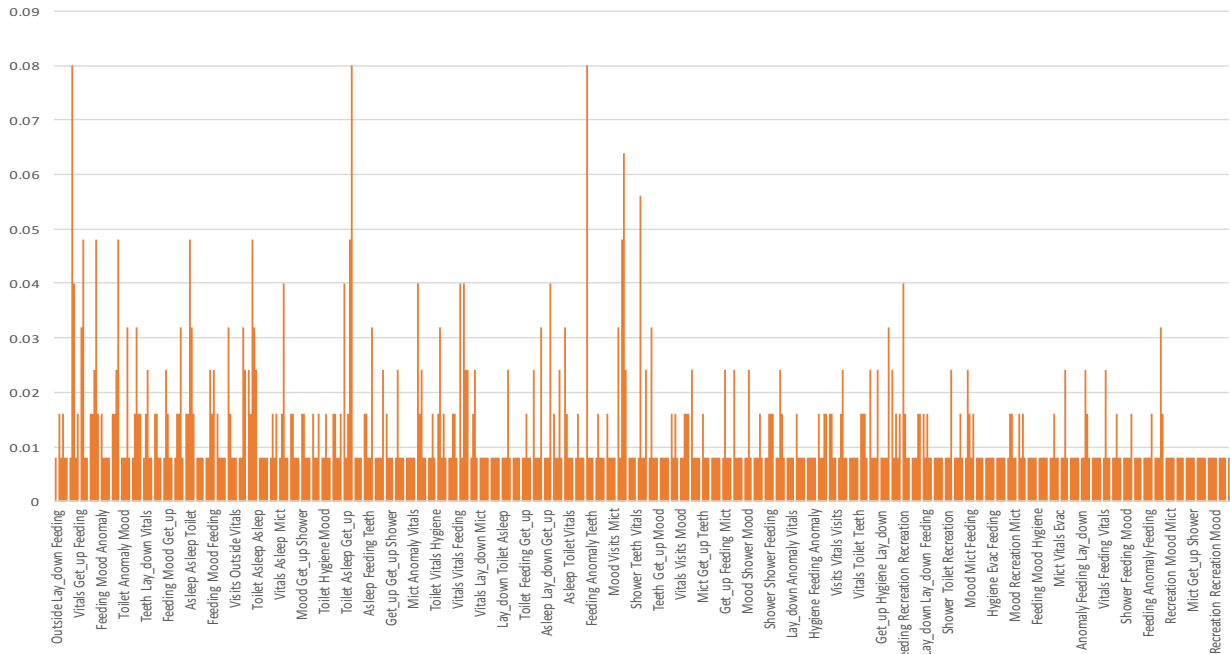


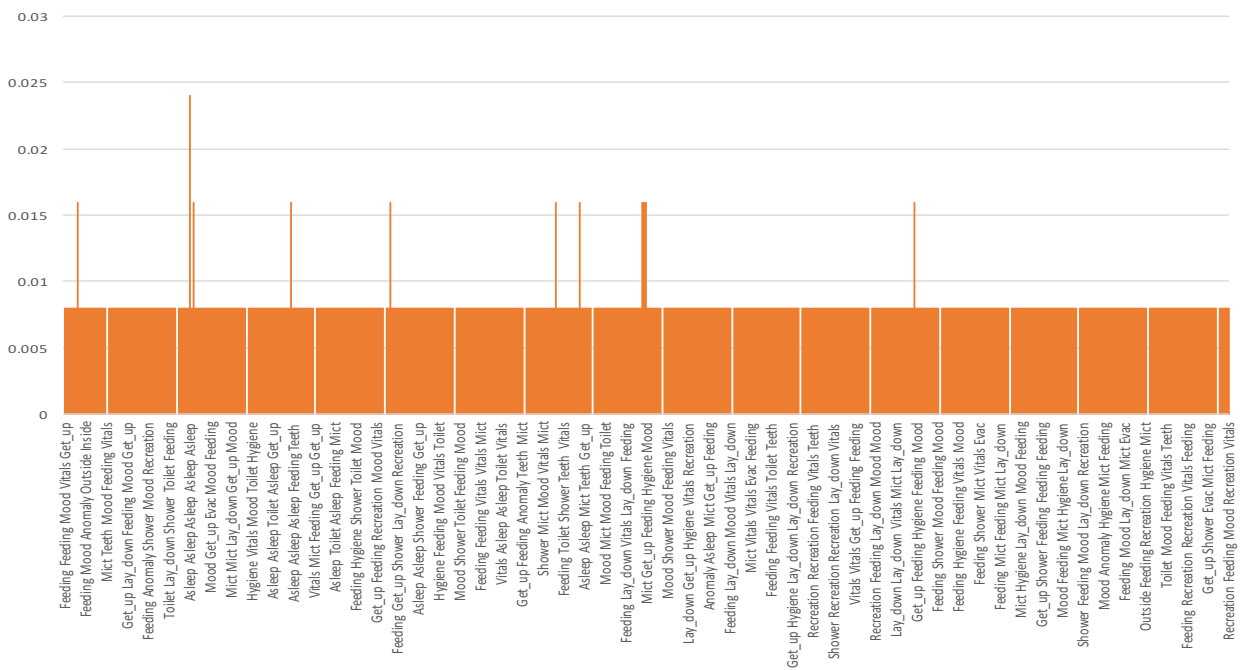
Figura 18. Longitud de la secuencia vs el número de patrones secuenciales descubiertos.

Resultados del análisis basado en secuencias. Adicionalmente al análisis basado en tiempo, se realizó un análisis basado en secuencia utilizando la extracción de patrones secuenciales como técnica. Con esta técnica se obtienen patrones de secuencias de longitud k . Para determinar una longitud de secuencia apropiada se realizó un análisis con respecto a la longitud de la secuencia y el número de patrones secuenciales posibles que se pudieran encontrar, mostrados en la Figura 18. El número máximo de secuencias diferentes (800) se alcanza con una longitud de secuencia de 7. La complejidad computacional del procesamiento aumenta con secuencias más grandes. Con base en los resultados de este experimento se consideró que las secuencias de longitud desde $k = 3$ y hasta $k = 5$ ofrecen un buen equilibrio entre la longitud de la

secuencia, la cantidad de patrones secuenciales encontrados y el tiempo de procesamiento.



(a) $k = 3$.



(b) $k = 5$.

Figura 19. Frecuencia de patrones secuenciales del Residente06 con diferente longitud.

En las Figuras 19(a) y 19(b) se puede ver el número de patrones secuenciales variando la longitud k . De esta forma se generó la lista de patrones secuenciales más

frecuentes con los datos de entrenamiento y posteriormente realizar la clasificación de los datos de prueba comparando las secuencias diarias nuevas con las secuencias más frecuentes de cada residente.

El número de días anormales detectados con el análisis basado en tiempo se redujo, después de efectuar el análisis basado en secuencias. Por lo que la precisión en la detección de anomalías aumentó. Los días anormales detectados con base en el análisis de tiempo se analizaron desde el punto de vista de secuencias, debido a que en ciertos casos el análisis basado en tiempo no es suficiente para la clasificación de anomalías; por ejemplo, cuando un retraso en las actividades debido a que el residente pasó más tiempo dormido y esto provoca que todas las actividades se realicen en un horario distinto al habitual. Lo que no necesariamente es un indicador de que algo anormal está sucediendo. Con el fin de descartar este tipo de anomalías mal clasificadas en el análisis basado en tiempo, es que se realiza el análisis basado en secuencias.

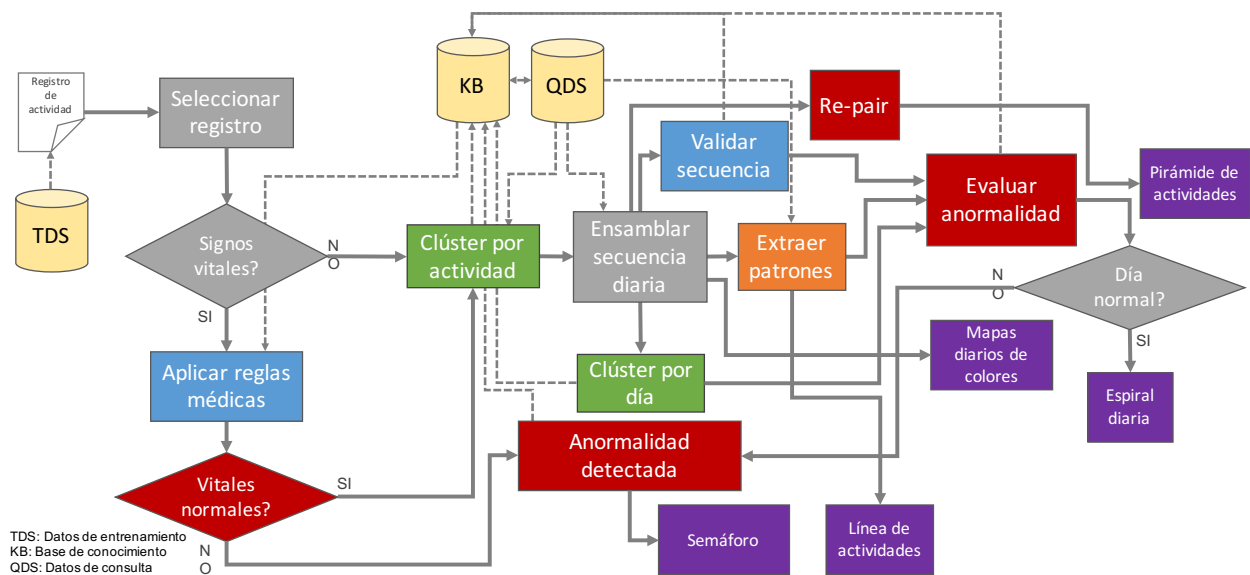


Figura 20. Diagrama de flujo de datos para la detección automática de anomalías empleando el componente híbrido.

Las anomalías detectadas se analizan también empleando las inferencias semánticas. En algunos casos es primordial esta evaluación, como en el caso de los registros de signos vitales. La Figura 20 muestra el diagrama de flujo de los datos donde se involucran todos los elementos del componente híbrido para la detección de anomalías en los registros de actividad de adultos mayores institucionalizados.

Los resultados finales de la clasificación se presentan en la Tabla 13.

Tabla 13. Detección automática de anomalías.

Residente	Número de días normales	Número de días anormales		
		Light	Medium	Severe
01	274	7	9	615
03	297	11	6	595
04	180	16	10	465
06	217	12	12	645
07	365	8	8	479
08	212	7	12	671
10	6	5	3	105
15	40	10	9	219
17	286	3	3	497
18	201	13	9	660
20	421	16	13	135
21	134	13	6	414
24	93	-	-	119
26	70	6	3	217
28	90	2	-	302
31	100	2	1	335
36	240	2	1	266
38	99	1	-	375
43	138	-	-	99
45	60	-	-	70

5.2. Evaluación humana de anomalías

El objetivo de este experimento es verificar si un usuario puede detectar anomalías usando las diferentes metáforas de visualización descritas en el capítulo 4. Para este fin, se hicieron dos sesiones con grupos de usuarios distintos: el primer grupo fueron un total de 11 estudiantes (7 hombres y 4 mujeres), en el segundo grupo fueron un total de 5 cuidadores (3 hombres y 2 mujeres).

El diseño experimental fue intra-sujetos, es decir, el mismo grupo de imágenes correspondientes a cada metáfora fue presentado a todos los participantes. El experimento consistió en dos partes: en la primera parte cada participante observaba un grupo de ocho imágenes de las cuales las primeras siete representaban días de un residente clasificados como *normales*; la última imagen (la octava) podría representar un día *normal* o *anormal*. De acuerdo con lo que el participante observaba en las primeras siete imágenes, éste decidía si la última imagen representaba un día *normal* o

anormal, sin ninguna otra información adicional (SE).

En la segunda parte, se le proporcionaron al participante las respuestas correctas sobre a cuál grupo correspondían cada una de las imágenes que se les habían presentado anteriormente. El participante observaba durante un minuto las imágenes correctamente clasificadas. Este proceso fue considerado como entrenamiento (CE). Posteriormente, se le presentó al participante un nuevo grupo de imágenes, como en la primera parte del experimento. Este procedimiento fue realizado con distintas imágenes correspondientes a cada metáfora de visualización.

Grupo de estudiantes. Los resultados fueron alentadores. La tasa de falsos negativos (FN) para la *Pirámide de actividades*, la *Línea de actividades* y los *Mapas diarios de colores* fueron respectivamente 18%, 11% y 27% antes del entrenamiento y 12%, 11% y 9% después del entrenamiento. Esto sugiere que los resultados mejorarían en un entorno real, dado que en este experimento los participantes no son cuidadores profesionales y el entrenamiento fue solo durante un minuto. La Tabla 14) contiene todos los detalles del experimento.

Tabla 14. Resultados de la evaluación de metáforas con estudiantes.

	Pirámide de actividades						Línea de actividades						Mapas diarios de colores					
	Sin entrenamiento (SE)			Con entrenamiento (CE)			Sin entrenamiento (SE)			Con entrenamiento (CE)			Sin entrenamiento (SE)			Con entrenamiento (CE)		
	VP	FP	FN	VP	FP	FN	VP	FP	FN	VP	FP	FN	VP	FP	FN	VP	FP	FN
Totales ¹	38	7	10	41	7	7	46	3	6	46	3	6	28	12	15	47	3	5

Este experimento validó las metáforas de visualización sugeridas para observar cambios en los comportamientos de los residentes usando como participantes personas que no están involucradas en el cuidado del adulto mayor; quedando pendiente verificar si con los cuidadores, los expertos en el cuidado del adulto mayor, como participantes se obtienen los mismos o incluso mejores resultados.

Grupo de cuidadores. Las tasas de falsos negativos (FN) para la *Pirámide de actividades*, la *Línea de actividades* y los *Mapas diarios de colores* fueron respectivamente

¹VP = Verdaderos positivos, verdaderos positivos y verdaderos negativos se consideraron; FP = Falsos positivos; FN = Falsos negativos.

32 %, 16 % y 32 % antes del entrenamiento y 4 %, 4 % y 12 % después del entrenamiento. La Tabla 15) contiene todos los detalles del experimento.

Tabla 15. Resultados de la evaluación de metáforas con cuidadores.

	Pirámide de actividades						Línea de actividades						Mapas diarios de colores					
	Sin entrenamiento (SE)			Con entrenamiento (CE)			Sin entrenamiento (SE)			Con entrenamiento (CE)			Sin entrenamiento (SE)			Con entrenamiento (CE)		
	VP	FP	FN	VP	FP	FN	VP	FP	FN	VP	FP	FN	VP	FP	FN	VP	FP	FN
Totales	14	3	8	21	3	1	20	1	4	21	3	1	14	3	8	19	3	3

En la Tabla 16 se muestran las medidas de precisión y recuerdo (*recall*) para los dos grupos de participantes en la evaluación. Estas medidas relacionan los valores de VP con FP (ecuación 2) y VP con FN (ecuación 3), respectivamente.

$$Precision = \frac{VP}{VP + FP} \quad (2)$$

$$Recuerdo = \frac{VP}{VP + FN} \quad (3)$$

Como se puede observar los resultados de precisión y recuerdo entre ambos grupos; y para cada una de las tres metáforas de visualización son muy similares. En todos los casos hubo un incremento en las mediciones después de que los participantes recibieron el entrenamiento con cada metáfora de visualización, excepto en la precisión de la línea de actividades del grupo de cuidadores. Esto da indicios sobre la utilidad de las metáforas de visualización para identificar anomalías en los datos presentados.

²SE = sin entrenamiento; CE = con entrenamiento.

Tabla 16. Medidas de precisión y recuerdo de la evaluación humana de anomalías.

Grupo	Medida	Pirámide de actividades ²		Línea de actividades		Mapas diarios de colores	
		SE	CE	SE	CE	SE	CE
Estudiantes	Precisión	0.8444	0.8542	0.9388	0.9388	0.7000	0.9400
	Recuerdo	0.7917	0.8572	0.8846	0.8846	0.6512	0.9038
Cuidadores	Precisión	0.8235	0.8750	0.9524	0.8750	0.8235	0.8636
	Recuerdo	0.6364	0.9545	0.8333	0.9545	0.6364	0.8636

5.3. Validación del componente híbrido

A continuación se contrasta la detección automática de anomalías basada en tiempo de 125 días etiquetados del Residente01; los resultados se muestran en la Figura 21. Se observa que el día 10 de octubre de 2012 fue clasificado como *severamente anormal*, esto se verificó con la revisión de otros registros de actividades alrededor de la fecha en cuestión, los cuales indicaron inconsistencias fuera del horario habitual en las actividades del residente.

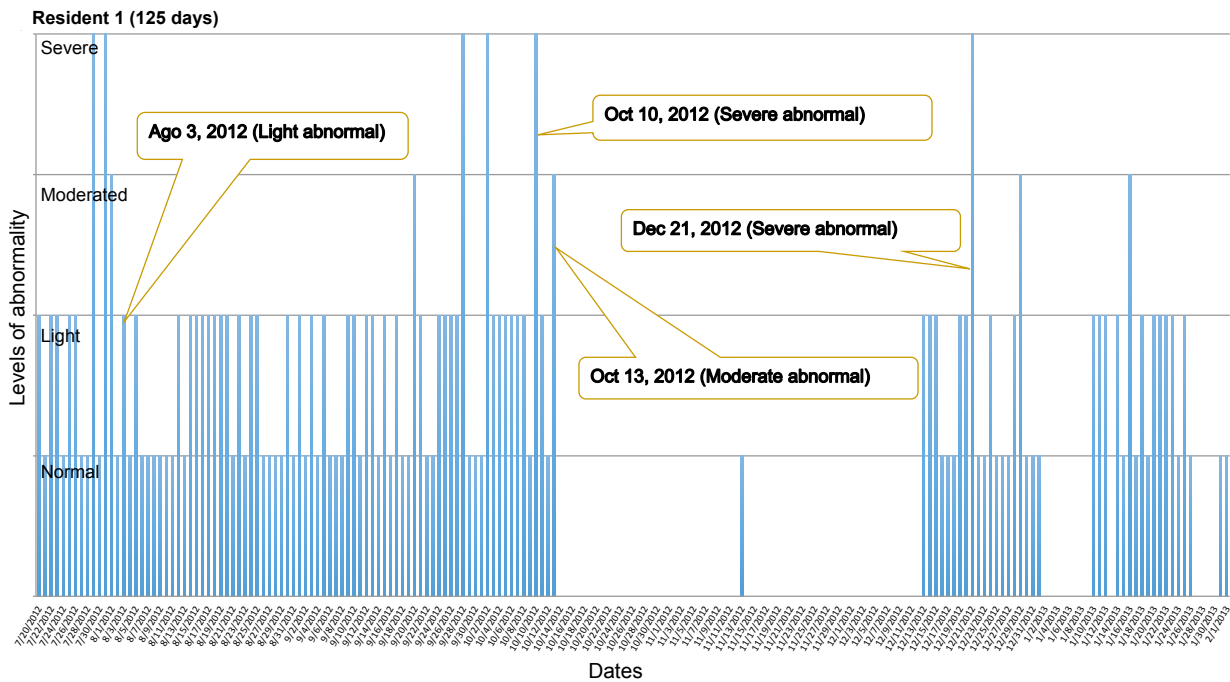


Figura 21. Validación del análisis basado en tiempo.

Por ejemplo, a las 12:58 el residente estaba acostado (*Lay down*) en su cama en vez de estar comiendo. Adicionalmente, el médico acudió a revisar en dos ocasiones al residente (*Medical attention*) en la misma semana, lo cual es extraño que ocurra porque las consultas médicas se realizan cada dos semanas o cada mes. También aparecieron nuevos registros de medicamentos (*Medication*), indicando un incremento en el número de medicamentos que consumía habitualmente el residente. Por último, se revisaron las anotaciones manuales adicionales de los cuidadores, a través de la regla de inferencia mostrada en la Tabla 17, donde indica que el residente fue sometido a una cirugía de ojos por el problema de cataratas que padecía. Todo lo anterior justifica y explica las visitas médicas y la ingesta de medicamentos oftálmicos (prednefrin y

gatifloxacino) después de esta fecha, lo que concuerda con las observaciones y anotaciones manuales realizadas por los cuidadores.

Tabla 17. Inferencias semánticas.

Clase de reglas	Ejemplos
a)	<p>Estados del residente</p> <p>[inferStatesFromActivities: (?r ps:is ps:Resident), (?c ps:is ps:Caregiver), (?a ps:is ps:Activity), (?r ps:is_caring_by ?c), (?c ps:registers ?a), (?a ps:has_description Activity:Get_up), noValue(?r ps:has_state State:Awake)] ->(?r ps:has_state State:Awake)</p> <p>Vitals(?vit), Record(?rec), Normal(?state), pulse(?vit, ?pv), greaterThan(?pv, 59), lessThanOrEqual(?pv, 100) ->hasState(?rec, ?state)</p> <p>Vitals(?vit), Record(?rec), LowAbnormal(?state), temperature(?vit, ?tv), lessThan(?tv, 36.5) ->hasState(?rec, ?state)</p> <p>Vitals(?vit), Record(?rec), Pre-highAbnormal(?state), lowBloodPressure(?vit, ?lbp), highBloodPressure(?vit, ?hbp), greaterThan(?lbp, 79), lessThanOrEqual(?lbp, 89), greaterThan(?hbp, 119), lessThanOrEqual(?hbp, 139) ->hasState(?rec, ?state)</p> <p>Vitals(?vit), Record(?rec), Abnormal(?state), pulse(?vit, ?pv), greaterThan(?pv, 100), lessThanOrEqual(?pv, 59) ->hasState(?rec, ?state), Vitals(?vit), Record(?rec), Abnormal(?state), breath(?vit, ?bv), greaterThan(?bv, 30), lessThanOrEqual(?bv, 10) ->hasState(?rec, ?state)</p> <p>Vitals(?vit), Record(?rec), LowAbnormal(?state), lowBloodPressure(?vit, ?lbp), highBloodPressure(?vit, ?hbp), lessThan(?hbp, 90), lessThan(?lbp, 60) ->hasState(?rec, ?state)</p>
b)	<p>Médicas</p> <p>Vitals(?vit), Record(?rec), Normal(?state), breath(?vit, ?bv), greaterThan(?bv, 10), lessThanOrEqual(?bv, 30) ->hasState(?rec, ?state)</p> <p>Vitals(?vit), Record(?rec), HighAbnormal(?state), lowBloodPressure(?vit, ?lbp), highBloodPressure(?vit, ?hbp), greaterThan(?hbp, 139), greaterThan(?lbp, 89) ->hasState(?rec, ?state)</p> <p>Vitals(?vit), Record(?rec), HighAbnormal(?state), temperature(?vit, ?tv), greaterThan(?tv, 37.2) ->hasState(?rec, ?state)</p> <p>Vitals(?vit), Record(?rec), Normal(?state), lowBloodPressure(?vit, ?lbp), highBloodPressure(?vit, ?hbp), greaterThan(?lbp, 59), lessThanOrEqual(?lbp, 79), greaterThan(?hbp, 89), lessThanOrEqual(?hbp, 119) ->hasState(?rec, ?state)</p> <p>Vitals(?vit), Record(?rec), Abnormal(?state), temperature(?vit, ?tv), greaterThan(?tv, 36.49), lessThanOrEqual(?tv, 37.2) ->hasState(?rec, ?state)</p> <p>(?r ps:is ps:Resident), (?r ps:food_intake percentage:50) ->(?r ps:food_behavior Intake:low)</p>
c)	<p>Comportamientos anormales</p> <p>(?r ps:is ps:Resident), (?r ps:hours_sleep quantity:10) ->(?r ps:sleep_behavior sleep:oversleep)</p> <p>(?r ps:is ps:Resident), (?r ps:has_mood Mood:sad), (?r ps:food_behavior Intake:low), (?r ps:sleep_behavior sleep:oversleep) ->(?r ps:has_behavior Behavior:abnormal)</p> <p>(?r ps:has_behavior Behavior:abnormal) ->(?r ps:visit_behavior ?x)</p>
d)	<p>Causas anormalidades</p> <p>(?r ps:has_behavior Behavior:abnormal) ->(?r ps:social_interactions_behavior ?x)</p> <p>(?r ps:has_behavior Behavior:abnormal) ->(?r ps:leisure_behavior ?x)</p>

En la Tabla 18 se observa el registro de actividades del día. Cabe mencionar que la inferencia no utiliza técnicas de procesamiento de lenguaje natural sobre información contenida en la descripción o del nombre de la medicina.

Tabla 18. Registro de actividades del día 10 de Octubre de 2012 del Residente01.

Id	Hora	Actividad	Descripción	Comentarios	Cuidador
14829	00:26	Medication	Tamsulosin		01
14834	08:20	Hygiene	Teeth		02
14835	08:21	Feeding	Breakfast		02
14836	08:22	Toilet	Evacuation and micturition		02
14837	08:23	Medication	Lasilacton		02
14838	08:24	Medication	Ferranina		02
14839	08:27	Medical Attention	Consultation	Nurse took him out	02
14904	12:56	Inside	Back	He returned from consultation	02
14905	12:57	Toilet	Micturition		02
14906	12:58	Check Over	Lay Down		02
14907	12:59	Hygiene	Hands		02
14908	13:05	Medical Attention	Special Treatment	He must rest due to his eyes surgery	02
14915	14:18	Feeding	Lunch		02
14922	14:20	Check Over	Lay down		02
14925	15:49	Medication	Aspirine, Protect		02
14926	15:50	Medication	Prednefrin		02
14928	16:37	Medication	Gatifloxacino		02
14930	17:48	Medication	Gatifloxacino		02
14931	17:49	Medication	Prednefrin		02
14940	18:27	Feeding	Dinner		02
14941	18:28	Toilet	Micturition		02
14944	18:32	Toilet	Micturition		02
14957	20:46	Medication	Gatifloxacino		01
14960	20:47	Medication	Dutasterida		01
14961	20:48	Medication	Prednefrin		01
14995	23:15	Toilet	Micturition		01

Otra validación de la clasificación se realizó a través de un muestreo de datos. Para lo cual se seleccionaron 48 días que fueron etiquetados manualmente para verificar si las clasificaciones fueron correctas o incorrectas. De los 48 días de muestra, 21 días fueron etiquetados manualmente como *normales*, de éstos el análisis basado en tiempo clasificó 14 como *normal*, 6 como *ligeramente anormal* y 1 como *medianamente anormal*. Para los días etiquetados manualmente como *anormales*, se clasificaron 7 como *normal*, 13 como *ligeramente anormal*, 1 como *medianamente anormal* y 6 como *severamente anormal*. Dependiendo del ajuste del umbral de normalidad (la lejanía de los puntos hacia cada centroide) se puede obtener un mayor o menor número de clasificaciones normales o anormales. Por ejemplo, para un umbral relajado, se puede tratar *ligeramente anormal* como *normal* obteniendo menos anomalías que en algún momento dado los cuidadores debieran atender; o por el contrario, siendo más

estricto con los umbrales, mayores anomalías detectadas se tendrán que atender.

Con el análisis basado en secuencias también se identificaron cambios dentro de periodos de tiempo mayores, por ejemplo por temporadas. En la Figura 22 se muestran los patrones de longitud $k = 4$ del Residente 06 en tres periodos diferentes de tiempo que concuerdan con las estaciones: verano, otoño y invierno. Las diferencias en las secuencias encontradas en los tres periodos corresponden a los cambios en el número de las interacciones sociales y de visitas, las cuales se incrementan debido a que se presentan las fiestas decembrinas en el invierno. Además, se observa que en los tres periodos hay patrones que permanecen constantes y no cambian; mientras que también emergen patrones nuevos distintos a los de otros periodos, exclusivos de un solo periodo.

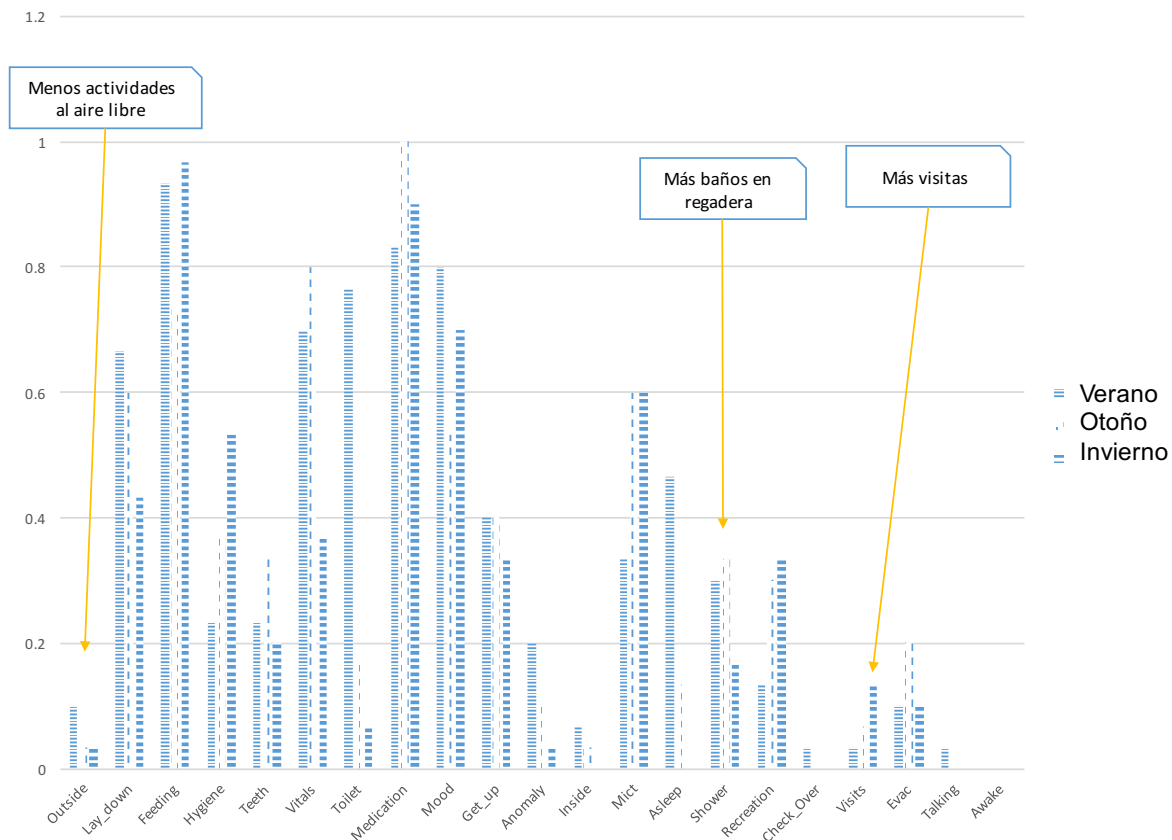


Figura 22. Validación del análisis basado en secuencias.

Los resultados anteriores indican que es posible identificar cambios en el comportamiento a través del tiempo utilizando las técnicas del componente híbrido. El componente híbrido no es una herramienta para el diagnóstico, es más bien una herramienta

para incrementar la consciencia del cuidado de los cuidadores hacia los residentes. A continuación se presenta una herramienta alternativa para asistir a cuidadores en el cuidado del adulto mayor. Esta herramienta incorpora un componente buscador de patrones al que se le puede incorporar el componente híbrido presentado en el capítulo 4. Cabe mencionar que esta herramienta alternativa fue desarrollada al inicio del trabajo de tesis, lo que permitió obtener elementos de diseño que nutrieron el diseño e implementación del componente híbrido.

5.4. Otras alternativas de desarrollo

En (Soto-Mendoza *et al.*, 2015b) se presenta el diseño de un calendario predictivo como una herramienta de asistencia para cuidadores en el cuidado del adulto mayor. A través del análisis de los registros de actividades se estudió la viabilidad de la implementación y se evaluó la propuesta de diseño con cuidadores en un entorno real. Esta herramienta está conformada de los siguientes componentes:

Tabla 19. Lista de estados de un residente dentro de la residencia geriátrica.

Estado	Definición	Medido por
Dormido (<i>Sleeping</i>)	El residente se encuentra en su recámara o en la sala y tiene un patrón de movimiento limitado o nulo. Típicamente este estado ocurre por la noche, pero también puede ocurrir durante la mañana o la tarde después de ingerir alimentos.	Inspección física por parte del cuidador, sensores de movimiento, presión y proximidad.
Comiendo (<i>Eating</i>)	El residente está comiendo por sí mismo o con ayuda de un cuidador. El residente realiza esta actividad en el comedor comunitario o en la privacidad de su recámara cuando no desea el acompañamiento de los demás residentes.	Inspección visual por parte del cuidador o cocinero, a través de la ubicación o de sensores de presión y proximidad.
Socializando (<i>Socializing</i>)	Ocurre cuando el residente interactúa (juega o platica) con otra persona.	Observación visual por el cuidador, sensores de proximidad o por reconocimiento de voz.
Despierto (<i>Awake</i>)	El residente acaba de despertarse y levanta de la cama.	Observación visual por el cuidador, sensores de movimiento, proximidad o presión (colocados debajo de la cama).
Al exterior (<i>Outside</i>)	El residente acude a realizar actividades al aire libre, fuera de la residencia. Salidas al cine, al teatro, salón de belleza, etc.	Inspección visual por el cuidador, ubicación o sensores de proximidad.
Aseado (<i>Cleaned</i>)	El residente se baña en regadera, el cuidador lo cambia de pañal, se lava las manos o se cepilla los dientes en el cuarto de baño.	Inspección visual por el cuidador, ubicación.
Medicando (<i>Medicated</i>)	El cuidador o enfermero le proporciona su medicamento al residente.	Anotaciones de cuidadores.
Entretenido (<i>Entertained</i>)	El residente participa en dinámicas de grupo juega un juego de mesa en el comedor, sala o jardín. También puede estar viendo televisión. Por lo regular es un grupo numeroso de residentes que participan y las actividades ocurren en horarios entre comidas.	Inspección visual del cuidador, ubicación o sensores de proximidad.

Pre-procesador de datos. Su tarea consiste en limpiar y dar formato a los registros de actividad. Se generó un modelo de actividad basado en estados donde se utilizó una máquina de estados finitos en la cual cada actividad fue caracterizada como un estado. Los estados que se definieron se muestran en la Tabla 19

Administrador de fusión de datos. Este componente discrimina o combina datos en función de su nivel de confiabilidad para mejorar la precisión de las inferencias o para completar los datos faltantes. Podemos observar, a partir de las definiciones anteriores, que los cuidadores o los diferentes tipos de sensores pueden recopilar información que se puede usar para inferir los estados de un residente. Sin embargo, existen grandes desafíos en la detección de medidas precisas. Por ejemplo, detectar una sesión de juegos recreativos con residentes y cuidadores que interactúan entre sí puede ser muy complicado para los sensores; cuando se encuentran en áreas comunes, la información recopilada proviene de varias fuentes.

Analizador de datos. Este componente toma la representación de estado del pre-procesador de datos y lo transforma en una matriz de transición de estados (ver Figura 23). Este arreglo, aunque simplista, es bastante informativo. Para mostrar su funcionalidad se procesaron los registros de 14 sujetos (residentes marcados como S1 a S14) para obtener las transiciones entre estados. Luego se contabilizó la cantidad de veces que un estado sigue a otro. En este proceso, los datos se dividieron en 8 horas, según la fecha y hora. El primer fragmento incluye actividades entre las 6 a.m. y 2 p.m., el segundo entre las 2 p.m. y 10 p.m. y el tercero a partir de las 10 p.m. a hasta las 6 a.m. del día siguiente. En general, los cambios de estado ocurrieron a menudo desde las 6 a.m. hasta las 2 p.m. porque los residentes y los cuidadores realizaron más actividades dentro de este horario (por ejemplo, tomar duchas, cambiarse de ropa, tomar el desayuno y tomar medicamentos al comienzo del día). Luego, se observó una disminución de las transiciones durante los siguientes dos cortes de tiempo (tarde y noche) (ver Figura 24).

Por ejemplo, el residente 14 (S14) tiene la mayor cantidad de transiciones en un segmento de tiempo dado. Éste también es el residente que tiene más necesidades de atención, porque padece la enfermedad de Parkinson en un estado avanzado. El sujeto

6 a.m. to 2 p.m.	Sleeping	Eating	Cleaned	Awake	Medicated	Entertained	Outside	Socializing
Sleeping	0	0	0	2	1	0	0	0
Eating	2	5	19	4	54	1	0	0
Cleaned	0	12	16	0	7	0	1	0
Awake	0	47	12	2	2	0	0	0
Medicated	1	19	32	18	235	0	0	0
Entertained	0	0	0	0	0	0	0	0
Outside	0	1	0	0	0	0	0	0
Socializing	0	0	0	0	0	0	0	0

Figura 23. Matriz de transición de estados del residente S7.

S14 no puede realizar casi ninguna actividad sin asistencia, ya que no puede moverse, caminar o incluso hablar. En el caso del sujeto S12, las transiciones en la tarde fueron más que en la mañana y en la noche; en este caso el residente tenía deterioro cognitivo y por lo general se despertaba más tarde que el resto de la población, y sus actividades comenzaban tarde, lo que implica un cambio en el periodo de actividad, coincidiendo con el resto de las actividades realizadas por la tarde. Cuando los residentes se fueron a dormir, el número de transiciones se redujo significativamente.

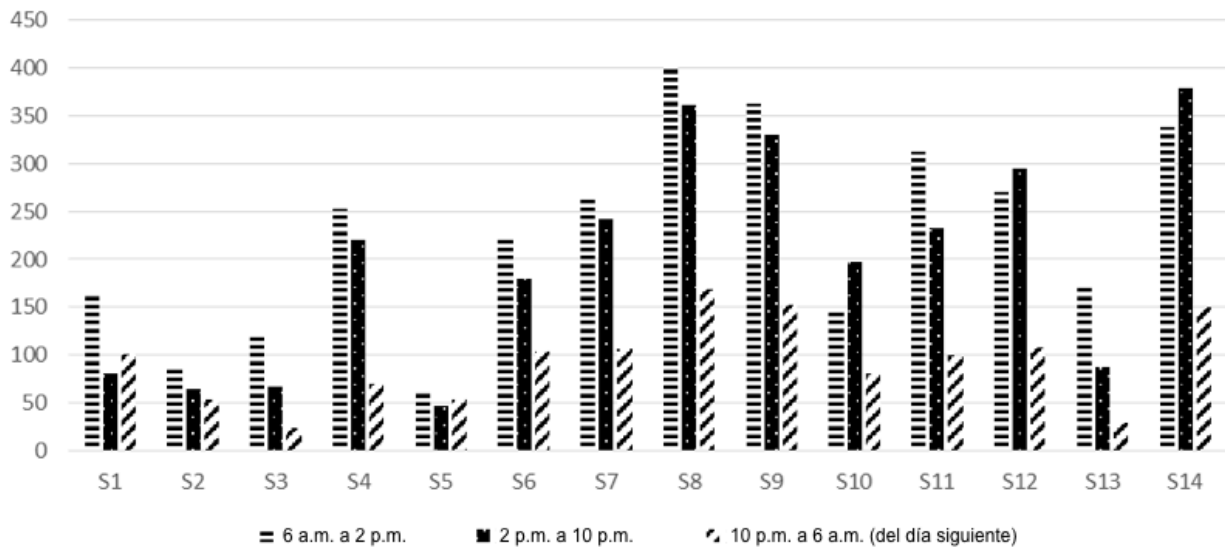


Figura 24. Número de transiciones de estado de cada residente durante la mañana, tarde y noche.

La matriz de transición de estados puede usarse para predecir el nivel de atención que necesita cada residente, lo cual es una información altamente valiosa para fines de planificación, ya que los residentes presentan diferentes condiciones de salud y,

por lo tanto, las necesidades de cada uno son variables. Otra aplicación para la matriz de transición de estados es separar a los residentes en grupos para asignar un cuidador más calificado para el grupo más exigente, es decir, el número de transiciones es un indicador muy consistente del nivel de independencia. En otras palabras, los residentes que todavía pueden hacer cosas por sí mismos tienen una menor cantidad de transiciones que los residentes que necesitan más ayuda.

Buscador de patrones. Este componente analiza el registro del estado de cada residente en busca de patrones o secuencias de estados dentro de un período de tiempo específico. Para esto se analizaron los estados realizados para cada residente entre semana y fines de semana. A partir de este análisis, se observa una diferencia en las transiciones entre estados. Durante los días entre semana el nivel de actividades es más alto comparado con los fines de semana: Esto sucede porque en algunos casos, la familia va de visita y se lleva a su residente a pasar un fin de semana fuera de la residencia. El buscador utiliza la matriz de transición para detectar días o períodos de tiempo habituales e inusuales para un residente determinado. La matriz también sirve como base para construir una narrativa útil para describir el día de un residente e informar a los familiares. Un ejemplo podría ser el siguiente:

*Durante los últimos 4 meses, las mañanas han sido habituales para Juanita. Ella comienza el día **dormida** y se **despierta** alrededor de las 7 a.m. Recibe su **medicina** y después de cambiarse de ropa, va al comedor a **desayunar**. Más tarde, el cuidador le da su **medicamento** y, luego, asiste a Juanita en lo que toma un **baño en regadera**. Cuando Juanita está **aseada**, da un **paseo por el jardín** y luego vuelve a **comer** porque tiene muy buen apetito todo el día.*

Las secciones en negrita en la narrativa anterior, provienen de la transición más probable en la matriz, después de un poco de pulido para la legibilidad. Para cada residente y cada transición, se agrega una frase. Se debe considerar que la matriz de transición de estados se elaboró manualmente.

Máquina de inferencias. También llamado motor de inferencias, utiliza la matriz de transición de estados y los patrones descubiertos para inferir el siguiente estado. Crea una generalización de máquinas de estados finitos para hacer predicciones. Los ejem-

plos más conocidos son los Modelos Ocultos de Markov (HMM, Hidden Markov Model) y los Procesos de Decisión de Markov Parcialmente Observables (POMDP, Partially Observable Markov Decision Process). La primera asume que el siguiente estado en una matriz de transición solo depende del estado actual, mientras que la posterior supone que no hay conocimiento sobre el estado actual del sistema y en su lugar mantiene una distribución de los estados actuales posibles del sistema. Los autómatas (mostrado en la Figura 25) que se construyó a partir de los datos son solo un ejemplo simple de ilustración. La parte más importante es definir los estados; una vez hecho esto, el proceso de aprendizaje es estándar.

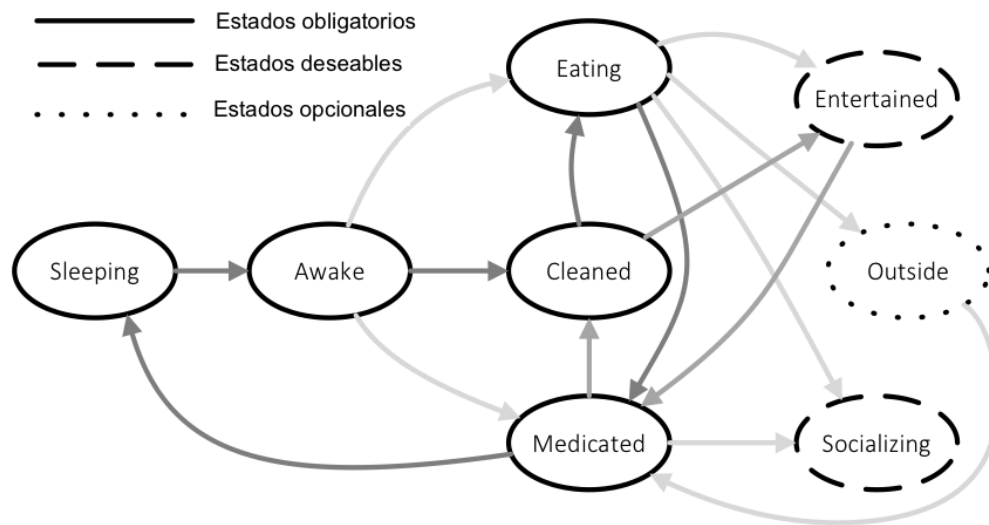


Figura 25. Ejemplo de una secuencia de estados siguiendo el camino más probable extraído de la matriz de transiciones.

Para esta propuesta, el predecir el próximo estado del residente con la mayor probabilidad es solo un aspecto del problema. El horario para un residente determinado puede o no considerar el estado previsto y, si se prevé, tendrá un grado de obligatoriedad. La información externa al modelo debe considerarse para el diseño. Es necesario establecer la obligatoriedad de un estado (que corresponde a una actividad). Otro parámetro sería asignar un umbral dinámico para cada residente. Al pasar el umbral se activará una notificación, por ejemplo.

Eventos no programados ni previstos por supuesto que pueden ocurrir. Piense, por ejemplo, en el residente que entra en la enfermería o en la cocina. Para estos eventos, existen técnicas ya publicadas (Bamis *et al.*, 2010) para evitar situaciones de riesgo para los residentes, por lo que no se consideran dentro del diseño del calendario como

un desafío.

La matriz de transición de estados es una forma de obtener información valiosa. Tiene la ventaja de estar en una forma muy legible y resulta ser una buena metáfora para visualizar la actividad de un residente como se ilustra en la Figura 26. Las secuencias de estados indican una trayectoria principal diaria de los estados de los residentes. Un día de actividad es una secuencia de estados que describe el día de un residente. Está lejos de adivinar cuál es el próximo estado más probable (con alta probabilidad). El módulo predictivo es la parte central del diseño del calendario predictivo. Los sensores y el personal humanitario son fuentes muy ricas de datos. Los algoritmos de reconocimiento de actividad no invasivos de fuentes de datos heterogéneos serán una buena fuente de contexto. Modelar a cada residente individualmente también es un desafío y podría proporcionar mejores predicciones.

6 a.m. - 2 p.m.	Sleeping	Eating	Cleaned	Awake	Medicated	Entertained	Outside	Socializing
Sleeping	0	0	0	2	1	0	0	0
Eating	2	5	19	4	54	1	0	0
Cleaned	0	12	16	0	7	0	1	0
Awake	0	47	12	2	2	0	0	0
Medicated	1	19	32	18	235	0	0	0
Entertained	0	0	0	0	0	0	0	0
Outside	0	1	0	0	0	0	0	0
Socializing	0	0	0	0	0	0	0	0

Figura 26. Camino individual siguiendo las transiciones entre estados.

Adicionalmente, se llevó a cabo una evaluación de uso y facilidad de uso del diseño final del calendario predictivo para validar los aspectos de diseño con los usuarios previstos. Los resultados se presentan a continuación.

5.4.1. Sesión de evaluación del diseño.

En la evaluación del diseño del calendario predictivo y su funcionalidad con los cuidadores se emplearon escenarios de uso y prototipos de baja fidelidad. Se organizó una sesión donde se presentó el diseño del sistema propuesto a un grupo de 4 cuidadoras y 2 personas en el área de enfermería (Tabla 20).

Tabla 20. Datos demográficos de los participantes en la sesión de diseño del calendario predictivo.

Identificador de participante	Edad	Sexo	Ocupación	Cargo
E1	32	Masculino	Enfermería	Coordinador de enfermería
C1	31	Femenino	Asistente de enfermería	Cuidadora
C2	-	Femenino	-	Cuidadora
C3	42	Femenino	-	Cuidadora
C4	47	Femenino	Profesora	Cuidadora
E2	22	Femenino	Enfermería	Enfermera

El calendario predictivo se presentó a los participantes y al final de la sesión se aplicó un cuestionario basado en el modelo de aceptación de tecnología (TAM) de 7 puntos (Davis, 1989) (ver resultados en la Tabla 21). En general, los cuidadores consideraron al calendario predictivo como una aplicación útil para su trabajo cotidiano siempre y cuando demande poca atención y reduzca el tiempo para el registro de actividades. Para los residentes, el uso de este tipo de aplicación fue percibido como menos útil (en comparación con los cuidadores) debido a la reticencia que los ancianos pueden expresar con la tecnología y las etapas avanzadas de sus condiciones médicas.

El cuestionario incluyó 22 aseveraciones relacionadas con la utilidad percibida y la facilidad de uso percibida. Por ejemplo, *"mi trabajo será más fácil con el calendario predictivo"* (PU) y *"Estoy interesado en usar el calendario predictivo"* (PEOU). Luego, se analizaron los comentarios de las sesiones y los cuestionarios, de donde se obtuvo evidencia cualitativa empírica de cómo los cuidadores percibirían el sistema al usarlo en su trabajo.

Los cuidadores evaluaron al calendario predictivo y estuvieron de acuerdo en que un sistema de administración de información anticipado sería útil. Mencionaron que incluir la detección de llanto es muy importante para detectar períodos de depresión. Además, algunos de los cuidadores estaban preocupados por la forma en que se identificaría la depresión, *"es un poco difícil saber el momento preciso en que alguien está deprimido, a menos que tenga un botón o algo así, pero es difícil porque en un momento los residentes están bien, y al siguiente no lo están, y eso es producto de su enfermedad"* [C2]. Una hipótesis que surge es que el análisis de los registros de las actividades y las anotaciones de los cuidadores podrían emplearse para identificar estas variaciones repentinas en el estado de ánimo. Entonces, el sistema podría proporcionar una buena predicción aproximada de los momentos en los que ocurrirán los

períodos de depresión.

Todos los cuidadores y enfermeras enfatizaron que el tiempo es un factor esencial. No tienen suficiente tiempo para registrar las actividades; por lo tanto, no usarán un sistema que los distraiga. Cuando se presentó el calendario predictivo, los cuidadores comentaron sobre la facilidad para registrar actividades porque todas las actividades ya estarían allí y solo los cuidadores las verificarían. Los cuidadores argumentaron que es muy complicado seguir una estricta secuencia de actividades debido al número de residentes y las diferentes tareas con cada uno de ellos, pero dijeron que el calendario predictivo les ayudaría a minimizar el tiempo para el registro de actividades.

La apariencia de la forma del calendario fue atractiva para los cuidadores porque quieren poder consultar accidentes previos o incidentes importantes sobre los residentes. Con el sistema actualmente en uso, no perciben la utilidad de registrar paso a paso todas las actividades porque no pueden consultar información almacenada previamente.

El calendario predictivo fue percibido como una herramienta para un "diagnóstico rápido de cada residente" debido al diseño de la vista individual de cada residente y la flexibilidad para cambiar el horario con las actividades de cada anciano. Se percibió como útil porque se presentaron en la misma pantalla y la navegación fue simple. Se acordó que el camino correcto es combinar diferentes fuentes de información contextual (por ejemplo, ubicación, movilidad, interacción social, datos médicos, signos vitales y actividades realizadas) para crear perfiles individualizados que permitan a los cuidadores conocer las necesidades de atención de cada residente. El conocimiento generado será digerido y utilizado por el calendario predictivo, que "aprenderá" de las anotaciones de los cuidadores, y administrará y organizará la información para que los cuidadores tengan información oportuna y específica sobre cada residente, sus hábitos y costumbres.

Los cuidadores destacaron las notificaciones de medicamentos, procedimientos especiales o signos vitales, como las características más importantes relacionadas con el cuidado de adultos mayores. La mayoría de los cuidadores y enfermeras prefieren una alarma o un modo de notificación de voz como recordatorio cuando se olvidan de realizar algunas actividades importantes, *"cuando por omisión olvidé hacer algún*

Tabla 21. Resultados del análisis de la aplicación del TAM.

Tipo de usuario	Criterio a evaluar	Número de pregunta	Media	Desviación estándar	Promedio media	Promedio desviación estándar
Cuidador	Utilidad percibida (PU)	Q1	5.83	1.47	5.60	1.36
		Q2	6.33	0.82		
		Q3	6.33	0.82		
		Q4	4.67	1.97		
		Q5	4.83	1.72		
	Facilidad de uso percibida (PEOU)	Q6	6.17	0.75	5.44	1.41
		Q7	5.67	1.03		
		Q8	4.83	1.72		
		Q9	5.50	1.64		
		Q10	5.00	1.67		
		Q11	5.50	1.64		
Adulto mayor	Utilidad percibida (PU)	Q12	3.83	1.72	4.39	2.19
		Q13	4.33	2.16		
		Q14	4.50	2.26		
		Q15	4.50	2.51		
		Q16	4.33	2.34		
	Facilidad de uso percibida (PEOU)	Q17	4.83	2.14	4.40	1.88
		Q18	4.67	2.07		
		Q19	4.50	2.07		
		Q20	4.83	1.47		
		Q21	4.00	1.67		
		Q22	4.00	2.10		

Nota: las preguntas fueron respondidas con base en la escala tipo Likert de siete puntos (1 = "Totalmente en desacuerdo", 7 = "Totalmente de acuerdo").

procedimiento en un momento específico, el sistema me lo recordará oportunamente" [C1].

En el sistema actualmente en uso, se detectó como un problema el tamaño del dispositivo utilizado. A veces, el tamaño del teléfono inteligente hace que sea imposible registrar las actividades debido a la falta de sueño del cuidador, la hora de la noche y la carga de trabajo durante el día. Como consecuencia, el tiempo consumido para el registro aumenta y los cuidadores pasan más tiempo después de que el turno ha concluido *"la realidad es que no hay suficiente tiempo"* [C3]. Se propuso el uso de tabletas con una pantalla y teclado más grandes, o computadoras; por lo tanto, el calendario predictivo debería ofrecer soporte para distintos dispositivos.

Las enfermeras se centraron más en la incorporación de otro tipo información médica al sistema; de esta forma, el diagnóstico, las decisiones y las acciones podrían centrarse más en el paciente, *"[el sistema] podría incorporar algunas actividades mé-*

dicas adicionales" [E1]. Las evaluaciones de enfermería se basan en los patrones que observaron en los residentes. Por lo tanto, el calendario predictivo debe considerar incorporar estos patrones de enfermería, "basamos nuestros diagnósticos en patrones. Cuando se obtiene un diagnóstico, proponemos un programa de atención priorizado" [E1]. Se podría incluir otra información médica para hacer que el sistema sea más útil para enfermeras, personal médico y gerentes de la residencia, "creo que sería una buena ayuda para las funciones de enfermería. En lugar de solo registrar las actividades, también implementaría el proceso de enfermería para apoyar legalmente a la compañía" [E2].

El calendario predictivo también fue percibido como una herramienta de entrenamiento para los nuevos cuidadores, *"Nosotros [los cuidadores] intentamos enseñarles [a los nuevos cuidadores] pero a veces somos muy egoístas y no enseñamos tan bien como deberíamos hacerlo" [C4]. Los nuevos cuidadores solo necesitan verificar y estudiar la información de la aplicación para saber qué hacer con cada residente y, a medida que pasa el tiempo, el nuevo cuidador aprenderá: "todos los datos deben registrarse en el sistema y el nuevo cuidador debe estudiar y repasar lo que debe hacer en su trabajo. En la práctica, será diferente, pero tendría [el cuidador nuevo] una noción previa" [C4]. Cabe mencionar que el calendario predictivo, no reemplazaría la capacitación personalizada y directa de un cuidador, pero sería un buen primer acercamiento al conocimiento general de cada residente y al proceso de cuidado de ancianos.*

5.5. Resumen

Este capítulo presenta los resultados de los experimentos realizados con el componente híbrido, la evaluación de las herramientas de visualización propuestas con cuidadores y la discusión correspondiente de los resultados.

Capítulo 6. Conclusiones, aportaciones y trabajo futuro

6.1. Conclusiones

En este trabajo de tesis se analizó un conjunto de datos provenientes de fuentes diversas relacionados con los adultos mayores institucionalizados. Dichas fuentes incluyen datos obtenidos a través de sensores y datos colectados a partir de las observaciones manuales por parte de los cuidadores. Con base en este análisis se obtuvo conocimiento sobre los tipos de datos y se diseñó, desarrolló y evaluó un componente híbrido para el análisis de comportamiento y la detección de anomalías de forma automática empleando técnicas guiadas por los datos y por el conocimiento.

Las preguntas de investigación establecidas al inicio de este trabajo de tesis quedaron respondidas de la siguiente manera:

[PI1] ¿Cuál debe ser la forma de representar los datos del repositorio de actividades de un adulto mayor que vive en una residencia geriátrica para la detección de patrones?

En la sección 3.3 se introdujo el registro de actividad como la estructura de datos para representar los datos del repositorio de actividades del adulto mayor institucionalizado. Es conocido que en la literatura existen otras formas de representar datos (Kim *et al.*, 2010; Aztiria *et al.*, 2012; Rashidi y Cook, 2013; Lago *et al.*, 2015); sin embargo, estas representaciones no incluyen varios tipos de datos como lo hace el registro de actividad. El registro de actividad, es la estructura que sirvió como base para la búsqueda y extracción de patrones en las actividades diarias considerando tanto datos obtenidos de sensores (como tradicionalmente se realiza en trabajos previos) así como datos de anotaciones manuales realizadas por los cuidadores.

[PI2] ¿Qué tipo de patrones de actividad se pueden encontrar mediante el análisis de las actividades de la vida diaria del adulto mayor que vive en una residencia geriátrica?

Los tipos de datos (sección 3.1) y el análisis preliminar del repositorio de datos (sección 3.2) fueron las bases para definir los tipos de patrones que se buscarían dentro

del repositorio. Estos patrones son complementarios entre sí para identificar anomalías, cada uno explota alguna característica del repositorio de datos. Los patrones son: los basados en tiempo, los basados en secuencia y los semánticos.

[PI3] ¿Qué tipo de técnicas de análisis se pueden emplear para detectar variaciones en las actividades del adulto mayor que vive en una residencia geriátrica?

Las técnicas de análisis empleadas para la detección de variaciones en las actividades del adulto mayor se describieron en la sección 4.3. El agrupamiento a través de clúster, la minería de patrones secuenciales, compresión de secuencias y las inferencias semánticas. Se sabe que existen otras técnicas que se han empleado en la detección de anomalías (Shin *et al.*, 2011; Mahmoud *et al.*, 2012; Zhao *et al.*, 2014; Barsocchi *et al.*, 2015), éstas analizan datos de sensores debidamente estructurados y homogenizados. Las técnicas empleadas en este trabajo se enfocaron en explotar las características encontradas en los datos del repositorio, los cuales son heterogéneos.

[PI4] ¿Cuáles son las formas más eficaces, desde la perspectiva del cuidador, para visualizar los datos de actividades del adulto mayor que vive en una residencia geriátrica?

En la sección 4.5 se muestran las distintas metáforas para la visualización de datos propuestas, todas ellas alimentadas por los resultados de aplicar las técnicas de análisis. En la sección 5.2 se presentaron los resultados de la evaluación por los cuidadores de las metáforas para visualizar las actividades del adulto mayor. A través de las medidas de precisión y recuerdo es posible obtener una medida cuantitativa de la eficacia, es decir, qué tanto ayudaron las metáforas a cumplir con su función. En este caso, para la correcta visualización e identificación de anomalías. De esta forma, la *pirámide de actividades* y la *línea de actividades* son las metáforas que obtuvieron los mejores resultados de precisión y recuerdo, por lo que resultaron ser, para los cuidadores, las más eficaces para la visualización e identificación de anomalías.

El pre-procesado de los datos es una tarea ardua y a la que se requiere invertir un

tiempo considerable para contar un repositorio de datos confiable y limpio que permita compartirlo para su reutilización dentro de la comunidad científica. A pesar de contar con herramientas (por ejemplo, *R*) que facilitan la labor de pre-procesamiento, además es necesario contar con un experto curador de datos, más aún cuando se tratan de datos naturalistas (como los utilizados en este trabajo de tesis); los expertos en los datos poseen un conocimiento específico de los tipos de datos, valores y rangos, lo cual impactaría directamente en el tiempo de pre-procesado y análisis de datos.

El análisis del comportamiento en adultos mayores se ha abordado con éxito relativo usando sensores. Sin embargo, no todas las instalaciones y las instituciones tienen acceso a este tipo de tecnología. Por otro lado, los registros anotados manualmente son ampliamente utilizados en instituciones geriátricas por razones legales y para responder a las consultas de familiares, médicos y gerentes. Este tipo de registros tienen una mayor cobertura que los datos del sensor y han recibido poca atención, ya que los datos anotados manualmente están en lenguaje natural. En este trabajo se siguió un enfoque diferente ya que se incorpora el uso de registros de actividad anotados manualmente junto con los datos obtenidos de sensores para identificar comportamientos anormales utilizando técnicas basadas en datos y basadas en conocimiento.

El componente híbrido detecta comportamientos anormales en las actividades diarias de adultos mayores. Se analizó la secuencia de actividades y el tiempo en que ocurrió cada una, pero no se consideró la duración de la actividad ni el tiempo de ocupación de sensores como lo incluyeron los trabajos previos de Dawadi *et al.* (2016) y Mahmoud *et al.* (2011), respectivamente. Además, en este trabajo se incorporaron varias fuentes de datos para predecir anomalías en comparación con otros trabajos donde se emplearon solo la ubicación (Rieping *et al.*, 2014), movilidad (Shin *et al.*, 2011), sensores (Mahmoud *et al.*, 2011) o actividades Dawadi *et al.* (2016).

6.2. Aportaciones

Las aportaciones de este trabajo de tesis son las siguientes:

- Un conjunto de herramientas y métodos para identificar patrones de actividad en adultos mayores institucionalizados analizando datos que provienen de sensores

y de anotaciones manuales por parte de los cuidadores.

- Un conjunto de metáforas para la visualización de los datos analizados para la identificación de anomalías en las actividades de adultos mayores institucionalizados.
- Evidencia empírica de la utilidad de las metáforas de visualización para identificar anomalías en las actividades de adultos mayores institucionalizados.
- El registro de actividad, el cual es una representación completa que incluye los datos de sensores y de anotaciones.
- Un repositorio de datos curado y longitudinal que contiene información de sensores (ubicación, signos vitales) y de las actividades diarias de adultos mayores. Este repositorio está disponible en [urlhttps://github.com/valeSM/HABITS.git](https://github.com/valeSM/HABITS.git).

Adicionalmente, de este trabajo de tesis se derivaron las siguientes publicaciones para la diseminación del mismo:

Artículos de revista arbitrada

- Valeria Soto-Mendoza, J. Antonio García-Macías, Edgar Chávez, Jorge R. Gomez-Montalvo, Eduardo Quintana, Detecting abnormal behaviours of institutionalized older adults through a hybrid-inference approach, *Pervasive and Mobile Computing*, Volume 40, 2017, Pages 708-723, ISSN 1574-1192, <http://dx.doi.org/10.1016/j.pmcj.2017.06.019>.
- Valeria Soto-Mendoza, J. Antonio García-Macías, Edgar Chávez, Ana I. Martínez-García, Jesús Favela, Patricia Serrano-Alvarado, and Maythé R. Zúñiga Rojas. 2015. Design of a Predictive Scheduling System to Improve Assisted Living Services for Elders. *ACM Trans. Intell. Syst. Technol.* 6, 4, Article 53 (July 2015), 31 pages. DOI=[urlhttp://dx.doi.org/10.1145/2736700](http://dx.doi.org/10.1145/2736700).

Talleres

- Tapia, A., Beltrán, J., Soto-Mendoza, V. and Caro, K., 2016, November. Designing visualization tools to support older adults care process. In *Computación Clínica e Informática Médica at Mexican International Conference on Computer Science*.
- Beltrán, J., Soto-Mendoza, V., García-Macías, J. A., & Caro, K. (2016, May). Smart spray-on clothing for disabled and older adults in AAL settings. In *Proceedings of the 10th EAI International Conference on Pervasive Computing Technologies for Healthcare* (pp. 308-312). ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).
- Caro, K., Beltrán, J., Martínez-García, A. I., & Soto-Mendoza, V. (2016, May). Exer-CaveRoom: a technological room for supporting gross and fine motor coordination of children with developmental disorders. In *Proceedings of the 10th EAI International Conference on Pervasive Computing Technologies for Healthcare* (pp. 313-317). ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).
- Soto-Mendoza, V., Beltrán, J., Chávez, E., Hernández, J., & García-Macías, J. A. (2015). Abnormal behavioral patterns detection from activity records of institutionalized older adults. *Human Behavior Understanding*, 119-131.
- Soto-Mendoza, V., Serrano-Alvarado, P., Desmontils, E., & Garcia-Macias, J. A. (2015, October). Policies composition based on data usage context. In *Sixth International Workshop on Consuming Linked Data (COLD2015) at ISWC*.
- Beltrán, J., Navarro, R., Chávez, E., Favela, J., Soto-Mendoza, V., & Ibarra, C. (2014, December). Detecting disruptive vocalizations for ambient assisted interventions for dementia. In *International Workshop on Ambient Assisted Living* (pp. 356-363). Springer, Cham.
- Soto-Mendoza, V., & Garcia-Macias, J. A. (2014, December). Lessons learned from a long-running assistive system for geriatric care. In *International Workshop on Ambient Assisted Living* (pp. 83-90). Springer, Cham. item Soto-Mendoza, V., Garcia-Macias, J. A., Martínez-García, A. I., Zúñiga-Rojas, M. R., Favela, J., & Serrano-Alvarado, P. (2013). Cuidadores virtuales para residentes reales. *Computación Clínica e Informática Médica*.

Carteles

- Zazueta, A. T., Márquez, J. B., & Mendoza, V. S. (2016). A visualization tool to support the decision making process of managers and caregivers from geriatric residences. *Avances en Interacción Humano-Computadora*, 1(1), 46-48.
- Soto-Mendoza, V. (2015). Identificación de patrones de comportamiento de adultos mayores que viven en residencias geriátricas. Taller sobre investigación en tecnologías de la información y envejecimiento. Ensenada, BC, México. Octubre 8-9.

Consortio doctoral

- Soto-Mendoza, V., (2014) "Discovery of activity patterns in Elders". 5th Workshop Uniting the Californias. Tijuana, BC, México. Junio 6-7.

6.3. Limitaciones

En este trabajo, los datos relevantes sobre el cuidado de adultos mayores se obtuvieron de forma no obtrusiva (a través de sensores y cuidadores), es decir, sin mucha intervención de los sujetos mismos. Aunque este paradigma es adecuado para el seguimiento y monitoreo de adultos mayores, los desarrollos futuros podrían considerar un enfoque más centrado en la persona (usuario) (Campbell *et al.*, 2008), donde la recolección de datos incluiría fuentes variadas y continuas desde teléfonos móviles, aplicaciones, sensores vestibles, etc. (Giordano y Puccinelli, 2015). Esto sin duda será un factor en la obtención de más tipos de datos. En este sentido, una limitación de este trabajo es la variedad del conjunto de datos utilizado para probar el enfoque propuesto. Es decir, los datos recopilados del repositorio están relacionados solo con las actividades diarias que realiza el adulto mayor dentro de una residencia geriátrica. Otros tipos de datos como datos de actividad física, GPS, calendarios, registro de vida y huellas digitales no fueron considerados por falta de infraestructura de sensado dentro de la residencia geriátrica.

Otra limitación de este trabajo está relacionada con la evaluación del enfoque. Esta evaluación se realizó a partir de datos históricos ya registrados y almacenados

previamente. Asimismo, para las evaluaciones de las metáforas de visualización el grupo de usuarios participantes fue reducido.

Las técnicas de análisis empleadas, si bien fueron elegidas con base en las características de los datos, no fueron contrastadas con otras técnicas debido a que el repositorio de datos carecía de una estructura homogénea, lo que dificultó la limpieza y curación de los datos. Esto provocó que una gran cantidad de tiempo fuera invertida en el pre-procesamiento de los datos.

6.4. Trabajo futuro

Se propone continuar realizando investigación con los resultados obtenidos en este trabajo de tesis en los siguientes aspectos:

Las técnicas de aprendizaje de automático, de minería de datos y de la web semántica demostraron ser en conjunto adecuadas para el descubrimiento de comportamientos anormales dentro de las actividades diarias de los adultos mayores. Este conjunto de técnicas puede ser enriquecido con la incorporación de otros métodos analíticos y reglas semánticas. Por otra parte, el enfoque híbrido propuesto en este trabajo puede incorporarse a plataformas del internet de las cosas (IoT, Internet of Things), como la presentada en Distefano *et al.* (2015), donde además de ofrecer el sensado de datos se incluya el análisis de los mismos detectando los patrones y las anomalías en los datos.

Los datos recopilados del repositorio están relacionados solo con las actividades diarias. Dentro de un entorno de IoT es posible recopilar datos precisos de dispositivos portátiles y teléfonos móviles (Qi *et al.*, 2016), sin perder de vista las implicaciones relacionadas con los adultos mayores y el uso de la tecnología (Detweiler y Hindriks, 2016). Los datos de actividad física, GPS, calendarios, registro de vida y huellas digitales sin duda enriquecerán y mejorarán las predicciones de nuestro enfoque. Sin embargo, los resultados presentados en este documento muestran que incluso cuando el conjunto de datos es limitado, es posible detectar anomalías dentro de las actividades diarias de los adultos mayores, y esto es útil dentro del paradigma "más con menos".

Los resultados obtenidos fueron a través del análisis de los datos de un repositorio

de una sola residencia geriátrica. Sin embargo, se pueden generalizar a otras residencias geriátricas. Basada en la experiencia empírica previa (Soto Mendoza, 2012) en tres diferentes residencias geriátricas, se constató que el tipo de anotaciones y la información relevante es similar entre ellas ya que el cuidado geriátrico es su tarea principal, y éste se rige por ciertos criterios que son estándar. El componente híbrido y las metáforas de visualización han sido propuestas siguiendo los requerimientos del cuidado geriátrico y principalmente de los cuidadores.

Por otro lado, este trabajo se puede generalizar y aplicar en otros entornos en donde se cuente con las siguientes características:

- Se lleve un registro de tareas o estados repetitivos, casi periódicos.
- Donde se pueden hacer observaciones objetivas y subjetivas.
- Donde las decisiones finales se toman por una persona, es decir, un algoritmo no es el que toma las decisiones de forma automática.
- Donde la carga de trabajo requiere que se aprecie con un simple vistazo, al juicio de la persona, si algún acontecimiento es usual o inusual.

Las características descritas anteriormente se pueden observar en cuestiones empresariales, en donde el gerente tiene un plan estratégico para la toma de decisiones que se nutre de información distinta proveniente de diversas áreas de la empresa (recursos humanos, producción, ventas, etc.); el sistema que apoya la toma de decisiones podrá emitir recomendaciones al tomador de decisiones, pero finalmente, éste basado en su experiencia emite su valoración y juicio para tomar la mejor decisión en el menor tiempo posible. Otra aplicación se puede dar en cuestiones de mercadotecnia, por ejemplo, en atención a clientes, evaluación de satisfacción al cliente o impacto publicitario. Por último, piense en un escenario educativo, en una escuela de atención para niños con capacidades diferentes más específicamente (Caro *et al.*, 2014), en donde tiene mucha importancia el avance de los niños y muchas de las tareas son repetitivas; las maestras y familiares proporcionan evaluaciones subjetivas y durante las terapias se evalúa objetivamente al niño. En la Figura 27 se muestran ejemplos de registros de una guardería y jardín de niños en donde los maestros y cuidadores registran información relevante sobre las actividades de los bebés y los niños. Análogamente al caso de

la atención de adultos mayores, el análisis de la información relacionada con el cuidado y bienestar de estos infantes puede beneficiar tanto al personal de la guardería como a los padres, al detectar de manera automática los cambios en los comportamientos Ye *et al.* (2015).

The image shows two pages of a handwritten activity log for a child in a nursery. The left page is dated 06-07-16 and includes sections for 'BAÑO', 'EJERCICIO DE ESTIMULACIÓN', and 'HORARIO DE SIESTA'. It has tables for tracking 'Limpio', 'Sucio', 'Cortas', 'Largas', 'Sucias', 'Rosado', and 'Normal' for 'UÑAS', and 'PAÑALERA' (Toallitas, C. Ropa) and 'ALIMENTACIÓN' (Desayuno, Comida, Merienda, Agua) with columns for 'MB', 'B', 'P', 'N'. The right page shows 'COMPORTAMIENTO' (BAÑO) with 'continuo' written, and 'OBSERVACIONES' with a note: 'Felicidad del [redacted] hizo papa en el baño. Pero un poco intento golpear a un compañero con el cepillo !!'. It also has a 'HORARIO DE SIESTA' section.

Figura 27. Ejemplo de los registros de actividades dentro de una guardería de niños.

Una evaluación *in situ* con usuarios reales podría proporcionar mediciones cualitativas sobre la utilidad de las predicciones y las metáforas de visualización. Esto se puede lograr a través de la implementación de prototipos de baja fidelidad para la evaluación de las metáforas de visualización siguiendo la metodología de evaluación de visualizaciones de datos propuesta por Mohamed y Ozkul (2016). Para esto se recomienda tener una muestra de participantes (cuidadores reales) en donde se puedan crear grupos ($n > 5$) y que cada grupo evalúe una metáfora distinta con el objetivo de comparar la utilidad y funcionalidad de cada una de las metáforas. Esto permitirá iterar el diseño y obtener una metáfora de visualización que se implemente en alguna residencia geriátrica para pruebas a largo plazo.

Se propuso una representación de los registros de actividad: uno explotando la agrupación por tiempo y la otra en cuanto a secuencias. El análisis de agrupación ha demostrado que la relación temporal entre las actividades es útil para determinar las desviaciones que solo dependen del tiempo. Para otro tipo de dependencias, el análisis de secuencias se puede usar para la detección de secuencias de actividad comunes.

Este tipo de análisis y técnicas podrían servir como puntos de entrada en sistemas de monitoreo más sofisticados para adultos mayores. La combinación de datos de los sensores y las anotaciones de los cuidadores podría conducir a mejorar los sistemas de cuidado y monitoreo de adultos mayores.

El objetivo de este trabajo es mostrar los resultados de un análisis personalizado de los residentes dentro de una residencia geriátrica. Sin embargo, este análisis puede extenderse al estudio los comportamientos de los grupos (Liang y Cao, 2015; Vu *et al.*, 2015). Por ejemplo, desde el punto de vista de un médico, la agrupación de adultos mayores con la misma enfermedad o condición crónica puede llevarse a cabo para descubrir tendencias similares en sus actividades diarias y para proporcionar un tratamiento oportuno. Además, desde la perspectiva comercial, se puede usar un análisis grupal para descubrir los grupos de adultos mayores que requieren más (o menos) cuidado y atención, para establecer un costo preciso y exacto de los servicios de atención.

Emparejar las actividades del repositorio con los patrones funcionales de salud (Gordon y Gordon, 1996; Gordon, 2003) con la finalidad de obtener una abstracción y simplificación de las actividades que brinde indicios de diagnósticos médicos para enfermeros y geriatras encargados del cuidado de adulto mayor. De esta forma se podría obtener un diseño de una metáfora de visualización que permita llevar un mejor control de los residentes y se busque un impacto con el personal del área médica.

Para finalizar, con las imágenes resultantes de la metáfora denominada *Pirámide de actividades* se puede generar un clasificador de imágenes para el análisis de la similitud entre éstas desde otra perspectiva. Contrastando los resultados con las clasificaciones obtenidas con las técnicas empleadas en este trabajo y con las evaluaciones humanas de anomalías.

Literatura citada

- Abu-Mostafa, Y. S., Magdon-Ismael, M., y Lin, H.-T. (2012). *Learning from data*, Vol. 4. AMLBook New York, NY, USA:.
- Agrawal, R. y Srikant, R. (1995). Mining sequential patterns. En: *Data Engineering, 1995. Proceedings of the Eleventh International Conference on*. IEEE, pp. 3–14.
- Archer, N., Fevrier-Thomas, U., Lokker, C., McKibbin, K. A., y Straus, S. E. (2011). Personal health records: a scoping review. *Journal of the American Medical Informatics Association*, **18**(4): 515–522.
- Aztiria, A., Augusto, J. C., Basagoiti, R., Izaguirre, A., y Cook, D. J. (2012). Discovering frequent user–environment interactions in intelligent environments. *Personal and Ubiquitous Computing*, **16**(1): 91–103.
- Bamis, A., Lymberopoulos, D., Teixeira, T., y Savvides, A. (2010). The behaviorscope framework for enabling ambient assisted living. *Personal and Ubiquitous Computing*, **14**(6): 473–487.
- Barsocchi, P., Cimino, M. G., Ferro, E., Lazzeri, A., Palumbo, F., y Vaglini, G. (2015). Monitoring elderly behavior via indoor position-based stigmergy. *Pervasive and Mobile Computing*, **23**: 26 – 42.
- Bergroth, L., Hakonen, H., y Raita, T. (2000a). A survey of longest common subsequence algorithms. En: *String Processing and Information Retrieval, 2000. SPIRE 2000. Proceedings. Seventh International Symposium on*. IEEE, pp. 39–48.
- Bergroth, L., Hakonen, H., y Raita, T. (2000b). A survey of longest common subsequence algorithms. En: *String Processing and Information Retrieval, 2000. SPIRE 2000. Proceedings. Seventh International Symposium on*. IEEE, pp. 39–48.
- Campbell, A. T., Eisenman, S. B., Lane, N. D., Miluzzo, E., Peterson, R. A., Lu, H., Zheng, X., Musolesi, M., Fodor, K., y Ahn, G.-S. (2008). The rise of people-centric sensing. *IEEE Internet Computing*, **12**(4): 12–21.
- Candás, J. L. C., Peláez, V., López, G., Ángel Fernández, M., Álvarez, E., y Díaz, G. (2014). An automatic data mining method to detect abnormal human behaviour using physical activity measurements. *Pervasive and Mobile Computing*, **15**: 228 – 241.
- Cao, L. y Philip, S. Y. (2012). *Behavior computing: modeling, analysis, mining and decision*. Springer.
- Caro, K., Martínez-García, A. I., Tentori, M., y Zavala-Ibarra, I. (2014). Designing exergames combining the use of fine and gross motor exercises to support self-care activities. En: *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility*. ACM, pp. 247–248.
- Chand, C., Thakkar, A., y Ganatra, A. (2012). Sequential pattern mining: Survey and current research challenges. *International Journal of Soft Computing and Engineering*, **2**(1): 185–193.
- Chen, H., Chiang, R. H., y Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS quarterly*, **36**(4): 1165–1188.

- Chernbumroong, S., Cang, S., Atkins, A., y Yu, H. (2013). Elderly activities recognition and classification for applications in assisted living. *Expert Systems with Applications*, **40**(5): 1662–1674.
- Chester, J. G. y Rudolph, J. L. (2011). Vital signs in older patients: age-related changes. *Journal of the American Medical Directors Association*, **12**(5): 337–343.
- Chiang, J.-H., Yang, P.-C., y Tu, H. (2014). Pattern analysis in daily physical activity data for personal health management. *Pervasive and Mobile Computing*, **13**: 13 – 25.
- Cohen, L. (2012). Time-frequency analysis. *Wavelets and Signal Processing*.
- Corno, F., De Russis, L., y Roffarello, A. M. (2016). A healthcare support system for assisted living facilities: An iot solution. En: *Computer Software and Applications Conference (COMPSAC), 2016 IEEE 40th Annual*. IEEE, Vol. 1, pp. 344–352.
- Daniel, K., Cason, C. L., y Ferrell, S. (2009). Assistive technologies for use in the home to prolong independence. En: *Proceedings of the 2nd International Conference on Pervasive Technologies Related to Assistive Environments*. ACM, p. 26.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, pp. 319–340.
- Dawadi, P. N., Cook, D. J., y Schmitter-Edgecombe, M. (2016). Modeling patterns of activities using activity curves. *Pervasive and Mobile Computing*, **28**: 51 – 68.
- Detweiler, C. y Hindriks, K. (2016). A survey of values, technologies and contexts in pervasive healthcare. *Pervasive and Mobile Computing*, **27**: 1 – 13.
- Distefano, S., Merlino, G., y Puliafito, A. (2015). A utility paradigm for iot: The sensing cloud. *Pervasive and Mobile Computing*, **20**: 127 – 144.
- Doucette, J. A., Khan, A., y Cohen, R. (2012). A comparative evaluation of an ontological medical decision support system (omed) for critical environments. En: *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*. ACM, pp. 703–708.
- Dusenbery, D. B. (2009). *Living at micro scale: the unexpected physics of being small*. Harvard University Press.
- Entrepreneur (2005). Get in on the hot senior care market. [Recuperado de <https://www.entrepreneur.com/article/80780> el día 1-Junio-2017].
- Fatima, I., Fahim, M., Lee, Y.-K., y Lee, S. (2013). A unified framework for activity recognition-based behavior analysis and action prediction in smart homes. *Sensors*, **13**(2): 2682–2699.
- Favela, J. (2013). Behavior-aware computing: applications and challenges. *IEEE Pervasive Computing*, **12**(3): 14–17.
- Fernández-Llatas, C., Garcia-Gomez, J. M., Vicente, J., Naranjo, J. C., Robles, M., Benedi, J. M., y Traver, V. (2011). Behaviour patterns detection for persuasive design in nursing homes to help dementia patients. En: *Engineering in medicine and biology society, EMBC, 2011 annual international conference of the IEEE*. IEEE, pp. 6413–6417.

- Gerrig, R. J., Zimbardo, P. G., Campbell, A. J., Cumming, S. R., y Wilkes, F. J. (2015). *Psychology and life*. Pearson Higher Education AU.
- Giordano, S. y Puccinelli, D. (2015). When sensing goes pervasive. *Pervasive and Mobile Computing*, **17, Part B**: 175 – 183.
- Gómez A, H. F., Martínez-Tomás, R., Tapia, S. A. A., y Zamorano, M. R. (2014). Using semantic technologies and the osu ontology for modelling context and activities in multi-sensory surveillance systems. *International Journal of Systems Science*, **45**(4): 798–809.
- Gordon, M. (2003). Manual de diagnósticos enfermeros.
- Gordon, M. y Gordon, M. (1996). *Diagnóstico enfermero*.
- Gröchenig, K. (2013). *Foundations of time-frequency analysis*. Springer Science & Business Media.
- Hartigan, J. A. y Hartigan, J. (1975). *Clustering algorithms*, Vol. 209. Wiley New York.
- Hartigan, J. A. y Wong, M. A. (1979). Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **28**(1): 100–108.
- Helaoui, R., Riboni, D., y Stuckenschmidt, H. (2013). A probabilistic ontological framework for the recognition of multilevel human activities. En: *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. ACM, pp. 345–354.
- Hoque, E. y Stankovic, J. (2012). Aalo: Activity recognition in smart homes using active learning in the presence of overlapped activities. En: *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2012 6th International Conference on*. IEEE, pp. 139–146.
- Jain, A. K. (2010). Data clustering: 50 years beyond k-means. *Pattern recognition letters*, **31**(8): 651–666.
- Joshi, J., Goecke, R., Alghowinem, S., Dhall, A., Wagner, M., Epps, J., Parker, G., y Breakspear, M. (2013). Multimodal assistive technologies for depression diagnosis and monitoring. *Journal on Multimodal User Interfaces*, **7**(3): 217–228.
- Kane, R. A. y Wilson, K. B. (2007). Improving practice through research in and about assisted living: implications for a research agenda. *The Gerontologist*, **47**(suppl_1): 4–7.
- Khan, A., Doucette, J. A., y Cohen, R. (2012a). A practical medical decision support system using structured knowledge & machine learning techniques.
- Khan, A., Doucette, J. A., Cohen, R., y Lizotte, D. J. (2012b). Integrating machine learning into a medical decision support system to address the problem of missing patient data. En: *Machine Learning and Applications (ICMLA), 2012 11th International Conference on*. IEEE, Vol. 1, pp. 454–457.
- Kim, E., Helal, S., y Cook, D. (2010). Human activity recognition and pattern discovery. *IEEE Pervasive Computing*, **9**(1).

- Lago, P., Jiménez-Guarín, C., y Roncancio, C. (2015). Contextualized behavior patterns for ambient assisted living. En: *Human Behavior Understanding*. Springer, pp. 132–145.
- Larsson, N. J. (1999). *Structures of string matching and data compression*. Lund University.
- Larsson, N. J. y Moffat, A. (2000). Off-line dictionary-based compression. *Proceedings of the IEEE*, **88**(11): 1722–1732.
- Le, T. (2014). Visualizing smart home and wellness data. *Handbook of Smart Homes, Health Care and Well-Being*.
- Levitis, D. A., Lidicker, W. Z., y Freund, G. (2009). Behavioural biologists do not agree on what constitutes behaviour. *Animal behaviour*, **78**(1): 103–110.
- Li, Z., Chu, C.-H., Yao, W., y Behr, R. A. (2010). Ontology-driven event detection and indexing in smart spaces. En: *Semantic Computing (ICSC), 2010 IEEE Fourth International Conference on*. IEEE, pp. 285–292.
- Liang, G. y Cao, J. (2015). Social context-aware middleware: A survey. *Pervasive and Mobile Computing*, **17, Part B**: 207 – 219.
- Lymberopoulos, D., Bamis, A., y Savvides, A. (2011). Extracting spatiotemporal human activity patterns in assisted living using a home sensor network. *Universal Access in the Information Society*, **10**(2): 125–138.
- Mahmoud, S. M., Lotfi, A., y Langensiepen, C. (2011). Abnormal behaviours identification for an elder's life activities using dissimilarity measurements. En: *Proceedings of the 4th International Conference on Pervasive Technologies Related to Assistive Environments*. ACM, p. 25.
- Mahmoud, S. M., Lotfi, A., y Langensiepen, C. (2012). User activities outlier detection system using principal component analysis and fuzzy rule-based system. En: *Proceedings of the 5th International Conference on Pervasive Technologies Related to Assistive Environments*. ACM, p. 26.
- Masseglia, F., Poncelet, P., y Teisseire, M. (2003). Incremental mining of sequential patterns in large databases. *Data & Knowledge Engineering*, **46**(1): 97–121.
- Merilahti, J., Pärkkä, J., y Korhonen, I. (2012). Estimating older people's physical functioning with automated health monitoring technologies at home: feature correlations and multivariate analysis. En: *Grid and pervasive computing workshops*. Springer, pp. 94–104.
- Minton, E. A. (2013). *Belief systems, religion, and behavioral economics: Marketing in multicultural environments*. Business Expert Press.
- Moffat, A. y Turpin, A. (2002). *Compression and coding algorithms*. Springer Science & Business Media.
- Mohamed, A. y Ozkul, T. (2016). User-interface usability evaluation. *International Journal of Computer Science and Security (IJCSS)*, **10**(2): 88.
- Mooney, C. H. y Roddick, J. F. (2013). Sequential pattern mining—approaches and algorithms. *ACM Computing Surveys (CSUR)*, **45**(2): 19.

- Moshtaghi, M., Zukerman, I., y Russell, R. A. (2015). Statistical models for unobtrusively detecting abnormal periods of inactivity in older adults. *User Modeling and User-Adapted Interaction*, **25**(3): 231–265.
- Ohmura, R. y Uchida, R. (2014). Exploring combinations of missing data complement for fault tolerant activity recognition. En: *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM, pp. 817–826.
- Okeyo, G., Chen, L., y Wang, H. (2014). Combining ontological and temporal formalisms for composite activity modelling and recognition in smart homes. *Future Generation Computer Systems*, **39**: 29–43.
- ONU (2015). Older person day. [Recuperado de <http://www.un.org/es/events/olderpersonsday/> el día 4-Noviembre-2015 @ 15:32].
- Orlando, S., Perego, R., y Silvestri, C. (2004). A new algorithm for gap constrained sequence mining. En: *Proceedings of the 2004 ACM symposium on Applied computing*. ACM, pp. 540–547.
- Oxford (2017). Behaviour. [Recuperado de <https://en.oxforddictionaries.com/definition/behaviour> el día 7-Agosto-2017].
- Patel, S. P., Shah, V., y Vala, J. (2015). Outlier detection in dataset using hybrid approach. *International Journal of Computer Applications*, **122**(8).
- Qi, J., Yang, P., Hanneghan, M., Fan, D., Deng, Z., y Dong, F. (2016). Ellipse fitting model for improving the effectiveness of life-logging physical activity measures in an internet of things environment. *IET Networks*, **5**(5): 107–113.
- Rao, V. C. S. y Sammulal, P. (2013). Survey on sequential pattern mining algorithms. *International Journal of Computer Applications*, **76**(12).
- Rashidi, P. y Cook, D. J. (2013). Com: A method for mining and monitoring human activity patterns in home-based health monitoring systems. *ACM Transactions on Intelligent Systems and Technology (TIST)*, **4**(4): 64.
- Rieping, K., Englebienne, G., y Kröse, B. (2014). Behavior analysis of elderly using topic models. *Pervasive and Mobile Computing*, **15**: 181 – 199.
- Robledo, L. M. G., Ortega, M. L., y Lopera, V. E. A. (2012). The state of elder care in mexico. *Current Translational Geriatrics and Experimental Gerontology Reports*, **1**(4): 183–189.
- Robledo, L. M. G., Campos, R. H. M., y Ortega, M. L. (2015). Present state of elder care in mexico. En: *Challenges of Latino Aging in the Americas*. Springer, pp. 379–392.
- Rodríguez, N. D., Cuéllar, M. P., Lilius, J., y Calvo-Flores, M. D. (2014). A survey on ontologies for human behavior recognition. *ACM Computing Surveys (CSUR)*, **46**(4): 43.
- Salomon, D. y Motta, G. (2010). *Handbook of data compression*. Springer Science & Business Media.

- Shin, J. H., Lee, B., y Park, K. S. (2011). Detection of abnormal living patterns for elderly living alone using support vector data description. *IEEE Transactions on Information Technology in Biomedicine*, **15**(3): 438–448.
- Shneiderman, B., Plaisant, C., y Hesse, B. W. (2013). Improving healthcare with interactive visualization. *Computer*, **46**(5): 58–66.
- Soto Mendoza, V. (2012). *Detección de situaciones de cuidado en adultos mayores institucionalizados*. Tesis de maestría, Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California.
- Soto-Mendoza, V. y Garcia-Macias, J. A. (2014). Lessons learned from a long-running assistive system for geriatric care. En: *International Workshop on Ambient Assisted Living*. Springer, pp. 83–90.
- Soto-Mendoza, V., Beltrán, J., Chávez, E., Hernández, J., y García-Macías, J. A. (2015a). Abnormal behavioral patterns detection from activity records of institutionalized older adults. En: *Human Behavior Understanding*. Springer, pp. 119–131.
- Soto-Mendoza, V., García-Macías, J. A., Chávez, E., Martínez-García, A. I., Favela, J., Serrano-Alvarado, P., y Rojas, M. R. Z. n. (2015b). Design of a predictive scheduling system to improve assisted living services for elders. *ACM Trans. Intell. Syst. Technol.*, **6**(4): 53:1–53:31.
- Soto-Mendoza, V., García-Macías, J. A., Chávez, E., Gomez-Montalvo, J. R., y Quintana, E. (2017). Detecting abnormal behaviours of institutionalized older adults through a hybrid-inference approach. *Pervasive and Mobile Computing*.
- Staab, S., Studer, R., Schnurr, H.-P., y Sure, Y. (2001). Knowledge processes and ontologies. *IEEE Intelligent systems*, **16**(1): 26–34.
- Stonefield, S. (2013). 16 negocios que puedes empezar mañana mismo. [[Recuperado de <https://www.forbes.com.mx/16-negocios-que-puedes-empezar-manana-mismo/> el día 16-Julio-2013 @ 11:00 am].
- Studer, R., Benjamins, V., y Fensel, D. (1998). Knowledge engineering: Principles and methods. *Data & Knowledge Engineering*, **25**(1): 161 – 197.
- Vu, L., Nguyen, P., Nahrstedt, K., y Richerzhagen, B. (2015). Characterizing and modeling people movement from mobile phone sensing traces. *Pervasive and Mobile Computing*, **17, Part B**: 220 – 235.
- Wagner, R. A. y Fischer, M. J. (1974). The string-to-string correction problem. *Journal of the ACM (JACM)*, **21**(1): 168–173.
- Wang, X. H., Zhang, D. Q., Gu, T., y Pung, H. K. (2004). Ontology based context modeling and reasoning using owl. En: *Proceedings of the Second IEEE Annual Conference on Pervasive Computing and Communications Workshops*, Washington, DC, USA. IEEE Computer Society, PERCOMW '04, pp. 18–.
- WHO (2015). Ageing. [Recuperado de <http://www.who.int/topics/ageing/es/> el día 4-Noviembre-2015 @ 15:26].

- Witten, I. H., Frank, E., Hall, M. A., y Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
- Xu, R. y Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions on neural networks*, **16**(3): 645–678.
- Yang, P., Stankevicius, D., Marozas, V., Deng, Z., Liu, E., Lukosevicius, A., Dong, F., Xu, L., y Min, G. (2016). Lifelogging data validation model for internet of things enabled personalized healthcare. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*.
- Ye, C., Xia, Y., Sun, Y., Wang, S., Yan, H., y Mehmood, R. (2015). Erar: An event-driven approach for real-time activity recognition. En: *2015 International Conference on Identification, Information, and Knowledge in the Internet of Things (IIKI)*. IEEE, pp. 288–293.
- Zhang, S., McCullagh, P., Nugent, C., y Zheng, H. (2011). An ontology-based context-aware approach for behaviour analysis. *Activity Recognition in Pervasive Intelligent Environments*, pp. 127–148.
- Zhao, Q. y Bhowmick, S. S. (2003). Sequential pattern mining: A survey. *ITechnical Report CAIS Nanyang Technological University Singapore*, pp. 1–26.
- Zhao, T., Ni, H., Zhou, X., Qiang, L., Zhang, D., y Yu, Z. (2014). Detecting abnormal patterns of daily activities for the elderly living alone. En: *HIS*. Springer, pp. 95–108.
- Zúñiga Rojas, M. R. (2012). *Cómputo móvil para el control y seguimiento de la salud de adultos mayores institucionalizados*. Tesis de maestría, Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California.

Anexos

Anexo A. Glosario

En esta sección se encuentra un listado con los términos más utilizados a lo largo de este trabajo de tesis, así como sus definiciones para facilitar el entendimiento del documento.

ABox. Es una forma de definir una ontología a través de aseveraciones, es decir, los hechos asociados al vocabulario de términos dentro de una base de conocimientos.

Comportamiento. Secuencia de actividades repetitiva, tomando en cuenta la frecuencia de la secuencia de actividades dentro de un periodo de tiempo específico.

CS1. Campaña de sensado número 1 donde se utilizaron los prototipos SSAMI y SMAMI para la colección de datos.

CS2. Campaña de sensado número 2 donde se utilizó el sistema Guardián para la colección de datos.

DDA. Data driven approach, perspectiva de análisis de comportamiento guiada por datos que utiliza técnicas de aprendizaje automático y minería de datos.

HDA. Hybrid driven approach, perspectiva híbrida de análisis de comportamiento que utiliza técnicas tanto de la DDA y de la KDA.

KB. Knowledge base, base de conocimientos

KDA. Knowledge driven approach, perspectiva de análisis de comportamiento guiada por conocimiento que emplea tecnologías semánticas.

OWL. Ontology Web Language, es un lenguaje para publicar y compartir datos en la web usando ontologías.

RDF/XML. Resource Description Format, es un lenguaje de propósito general para representar información con tecnologías semánticas en conjunto con la serialización del eXtended Markup Language (XML).

TBox. Es una forma de definir una ontología a través de sentencias que describen una conceptualización, un conjunto de conceptos y las propiedades asociadas a esos conceptos.

Anexo B. Resumen de técnicas guiadas por datos

A continuación encontrará los detalles del funcionamiento de las técnicas de minería de datos y aprendizaje automático que se utilizaron para el desarrollo del trabajo de tesis.

Clúster

La técnica de clúster, mejor conocido como *clustering* en inglés, es un algoritmo de agrupamiento de un grupo de datos siguiendo un criterio específico. Este criterio por lo general es la distancia o similitud entre los puntos de datos. Clúster es una técnica de clasificación no supervisada, es decir, es un análisis exploratorio donde no se cuenta con información *a priori*, no están disponibles los datos etiquetados en sus respectivas categorías (Hartigan y Hartigan, 1975).

Los algoritmos de clúster dividen los datos en un cierto número de grupos o *clusters*. Un clúster se define considerando la homogeneidad interna del grupo y la separación externa con respecto a otros grupos. Esto es, los patrones en el mismo grupo deben ser similares entre sí, mientras que los patrones entre diferentes grupos deben ser disimilares. Tanto la similitud como la disimilitud deben ser examinadas de forma clara y significativa (Xu y Wunsch, 2005).

El algoritmo de clúster K-medias es el siguiente (Jain, 2010). Sea $X = x_i, i = 1, \dots, n$ el conjunto de n puntos con d dimensiones a agrupar en K clusters, $C = c_k, k = 1, \dots, K$. K-medias encuentra una partición tal que el error cuadrático entre la media empírica de un clúster y los puntos dentro del clúster es mínimo. Sea μ_k la media del clúster c_k . El error cuadrático entre μ_k y los puntos dentro del clúster c_k está definido como $J(c_k) = \sum_{x_i \in c_k} \|x_i - \mu_k\|^2$. El objetivo de K-medias es minimizar la suma del error cuadrático de todos los K clusters, $J(C) = \sum_{k=1}^K \sum_{x_i \in c_k} \|x_i - \mu_k\|^2$.

Como se mencionó anteriormente, esta técnica se utilizó para determinar los tiempos habituales de ocurrencia de las diversas actividades de los residentes durante el día.

Extracción de patrones secuenciales

La extracción de patrones secuenciales es un método importante de minería de datos para determinar el comportamiento en bases de datos de secuencias a través del tiempo (Rao y Sammulal, 2013; Mooney y Roddick, 2013). Se define comúnmente como la búsqueda del conjunto completo de subsecuencias frecuentes en un conjunto de secuencias (Agrawal y Srikant, 1995). El patrón secuencial es una secuencia de conjuntos de elementos que ocurren con frecuencia en un orden específico. Todos los elementos en el mismo conjunto de elementos se supone que tienen el mismo valor de tiempo de transacción o que están dentro de un intervalo de tiempo. Cada secuencia corresponde a una lista de eventos ordenada cronológicamente, donde cada evento es una colección de elementos (conjunto de elementos) que ocurren simultáneamente. El orden temporal entre los eventos es inducido por las marcas de tiempo absolutas asociadas con los eventos (Orlando *et al.*, 2004).

Para el caso de las actividades de un adulto mayor, una transacción son las actividades (una o más) que se registran de un adulto mayor en un instante de tiempo determinado (hora y minuto) del día. El día es una secuencia ordenada cronológicamente de transacciones y varios días componen el repositorio de actividades del adulto mayor (Zhao y Bhowmick, 2003). El proceso de extracción de patrones secuenciales del repositorio de actividades del adulto mayor, es el siguiente:

Sea D el repositorio de actividades del adulto mayor compuesto por transacciones, donde cada transacción T tiene asociada un identificador, una marca de tiempo y un conjunto de actividades. Sea $A = a_1, a_2, \dots, a_3$ el conjunto de símbolos que representan a las actividades. El conjunto de actividades no puede ser un conjunto vacío. Una secuencia S es un grupo de conjuntos de actividades ordenados cronológicamente denotada como $\langle s_1, s_2, \dots, s_n \rangle$, donde $s_j, j \in 1, \dots, n$ corresponde a una actividad. Una secuencia- k es una secuencia de k actividades (o de longitud k). Una secuencia $\langle s_1, s_2, \dots, s_n \rangle$ es una subsecuencia de otra secuencia $\langle s'_1, s'_2, \dots, s'_m \rangle$ si existen enteros $i_1 < i_2 < \dots < i_n$ tal que $s_1 \subseteq s'_{i_1}, s_2 \subseteq s'_{i_2}, \dots, s_n \subseteq s'_{i_n}$ (Massegli *et al.*, 2003). El problema de la extracción de patrones secuenciales es encontrar todas las secuencias s tal que $supp(s) \geq minsup$ para una base de datos D , dado un umbral de soporte para $minsup$.

Subsecuencia común más larga

La subsecuencia común más larga es un método para la comparación de cadenas que consiste en encontrar el número máximo de símbolos comunes entre sí, preservando el orden de aparición de los símbolos (Bergroth *et al.*, 2000b). Formalmente, se comparan dos cadenas de símbolos, $X[1...m]$ y $Y[1...n]$, las cuales son elementos del conjunto Σ^* ; Σ denota el alfabeto que contiene σ símbolos. Una subsecuencia $S[1...s]$ de $X[1...m]$ se obtiene eliminando $m - s$ símbolos de X . Una subsecuencia común (CS) de $X[1...m]$ y $Y[1...n]$, $CS(X, Y)$, es una subsecuencia que aparece en ambas cadenas. La subsecuencia común más larga (LCS) de X y Y , $LCS(X, Y)$, es la subsecuencia común de longitud máxima. La longitud de $LCS(X, Y)$ está representada por $r(X, Y)$.

Este algoritmo fue utilizado en este trabajo de tesis para comparar la secuencia de actividades con el conjunto de las secuencias de actividades más frecuentes obtenidas a través del método extracción de patrones secuenciales, descrito anteriormente. De esta manera se determinó que tan similar era la secuencia de actividades y se estableció el nivel de anormalidad de la secuencia.

La forma tradicional de resolver el $LCS(X[1...m], Y[1...n])$ es determinando la subsecuencia común más larga para todas las combinaciones posibles de símbolos dentro de la cadenas a comparar. La recurrencia para obtener la longitud LCS para cada par de símbolos $X[1...i], Y[1...j]$ está dada por (Wagner y Fischer, 1974):

$$LCS(X_i, Y_j) = \begin{cases} \emptyset & \text{si } i = 0 \text{ y } j = 0 \\ LCS(X_{i-1}, Y_{j-1}) \cup x_i & \text{si } x_i = y_j \\ \max(LCS(X_i, Y_{j-1}), LCS(X_{i-1}, Y_j)) & \text{si } x_i \neq y_j \end{cases} \quad (4)$$

La solución para la subsecuencia común más larga entre X_i y Y_j se obtiene al comparar los elementos x_i y y_j . Si son iguales, entonces la secuencia $LCS(X_{i-1}, Y_{j-1})$ será extendida por el elementos x_i . Si no son iguales, entonces el más largo entre $LCS(X_i, Y_{j-1})$, y $LCS(X_{i-1}, Y_j)$, será la respuesta (Si ambos elementos son del mismo tamaño pero no son idénticos, ambas soluciones serán resultados).

Emparejamiento recursivo

El emparejamiento recursivo (*re-pair*), o *recursive pairing* en inglés, es un esquema de compresión que combina un método de derivación de frases y una compacta codificación de diccionario. El esquema es eficiente, particularmente en la descompresión, y tiene características que hacen que sea una opción favorable cuando se buscan directamente datos comprimidos. Este esquema de compresión forma parte de los esquemas conocidos como *off-line*, o fuera de línea, cuando se infiere el diccionario y durante la descompresión, se leen los bits y las frases se escriben de manera intercalada. Además, durante la decodificación, solo se almacena una representación compacta del diccionario. Por lo tanto, durante la descompresión, *re-pair* tiene una ventaja espacial contra los esquemas basados en diccionarios incrementales y los modelos basados en contexto (Larsson y Moffat, 2000).

El objetivo del modelo basado en diccionario es derivar un conjunto de frases (normalmente, pero no siempre, subcadenas parte del mensaje a codificar) de tal manera que reemplazando las ocurrencias de esas frases en el mensaje con referencias a la tabla de frases disminuye la longitud del mensaje. Adicionalmente, como la idea es incluir la tabla de frases como parte del mensaje comprimido, el esquema de derivación utilizado debe permitir una codificación compacta del conjunto de frases. Este último requisito no aplica a los métodos basados en diccionarios incrementales y pueden crear su diccionario sin preocuparse por cómo podría ser representado (Salomon y Motta, 2010).

El algoritmo de derivación de frases que utiliza *re-pair* consiste en reemplazar el par de símbolos más frecuente en el mensaje origen por un nuevo símbolo, reevaluando las frecuencias de todos los pares de símbolos con respecto al alfabeto ahora extendido y luego repitiendo el proceso hasta que no haya ningún par de símbolos adyacentes que aparezca dos veces (Moffat y Turpin, 2002).

El algoritmo es el siguiente:

1. Identificar los símbolos a y b tal que ab es el par de símbolos adyacentes más frecuente en el mensaje. Si no existe un par que aparezca más de una vez, detenerse.

2. Introducir un nuevo símbolo A y reemplazar todas las ocurrencias ab con A .
3. Repetir desde el paso 1.

Esta técnica de compresión fue utilizada para compactar las secuencias de actividades y reducir su representación para la búsqueda de patrones normales y anormales.