

Inteligencia Computacional en Sistemas de Teleasistencia en Domicilios

Thesis dissertation

Aitor Moreno Fdez. de Leceta

Departamento de Ciencias de la Computación e Inteligencia Artificial
Universidad del País Vasco



Directores: Prof. Manuel Graña. Dr. Jose Manuel Lopez-Guede

Junio, 2018

Resumen

Trabajo basado en un proyecto empresarial en un entorno productivo real:

- 1 Objetivo.
Implantar un sistema de detección de eventos de riesgo en el comportamiento habitual de personas ancianas que viven solas en sus domicilios.
- 2 Reto.
Problema no resoluble con técnicas de programación clásica.
Necesidad de aplicación de Técnicas de Inteligencia Artificial sobre un contexto real.

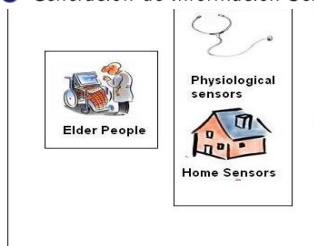
Aportaciones

- 1 Sistema de Generación de Resúmenes Clínicos.
- 2 Sistema combinado de razonadores basados en técnicas de Inteligencia Artificial:
 - Reglas Semánticas de Proceso.
 - Reglas Heurísticas introducidas por expertos en el dominio.
 - Módulo de Detección Automática de Patrones.

Contenidos

- 1 Introducción.
- 2 Generación de Resúmenes Clínicos.
- 3 Generación de Información Sensórica.
- 4 Implementación del Sistema.
- 5 Resultados de la Tesis.
- 6 Conclusiones y Trabajos Futuros.

3 Generación de Información Sensórica.

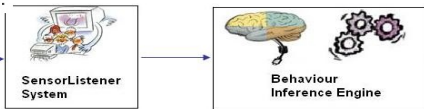


Elderly People Care at Home

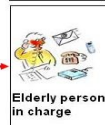
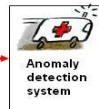
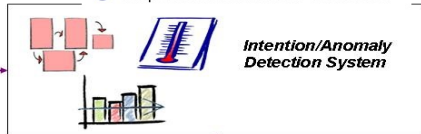
2 Generación de Resúmenes Clínicos.

Expert Knowledge

Relevant clinical record evidences into the teleassistance system



4 Implementación del Sistema.



5 Resultados del Proyecto.

Introducción.

- 1 Motivación.
- 2 Estado del Arte.
- 3 Esquema Funcional.

Introducción.

- 1 **Motivación.**
- 2 Estado del Arte.
- 3 Esquema Funcional.

Motivación

El envejecimiento en Europa se está incrementando rápidamente :

- Las personas mayores de 65 años se han incrementado desde un 17.4% en el 2010 a una estimación del 28.8% en el 2050.
- En el 2025 una de cada cuatro personas tendrá más de 65 años, y la mitad serán mayores de 50 años.

EL incremento en la demanda de soluciones de sistemas de teleasistencia:

- La preferencia de los europeos a envejecer en su propio hogar (64.4%).
- El incremento de las políticas públicas en la atención domiciliaria.
- El incremento de la capacitación de profesionales e instituciones.

Retos Actuales (I)

Los sistemas AAL (Ambient Assisted Living) necesitan superar algunos retos a día de hoy:

- Sensórica.
- Robótica Asistiva.
- Seguridad, Privacidad, Legalidad y Ética.
- Algoritmia: Sistemas actuales poco confiables y precisos.
- Personalización de patrones por domicilio.

Retos Actuales (II)

- 1 La “ETL” debe ser parametrizable, sencilla y directa:
 - Los sensores se conectan a la red de forma “Plug & Play”.
 - Integrados directamente en la Ontología del Sistema .
 - Reglas de Proceso semánticas.
- 2 Integrar Conocimiento de Expertos de forma usable:
 - Expertos en Teleasistencia.
 - Personal Clínico.
 - Reglas de Negocio (Reglas Heurísticas).

Retos Actuales (III)

Modelar y detectar de forma autónoma:

- Estado Clínicos del Usuario.
- Hábitos de los Usuarios.
- Variaciones sobre dichos Hábitos.
- Desviaciones sobre sus Patrones Habituales.

Se debe de proveer:

- Estado del Usuario.
- Evolución del Usuario.
- Descriptivo de la Alerta Generada.
- Recomendaciones de Acción, tanto Médicas como Asistenciales.

Proyectos empresariales

El trabajo de la presente tesis se ha desarrollado en el contexto de varios proyectos empresariales:

- “Chiron”: proyecto europeo.
- “HOI”: Etorgai liderado por Onkologikoa.
- “HousGai”: Etorgai liderado por Matia.
- “Ebizi”: Etorgai de liderado por Auzolagun.
- “Guarantee”: ITEA liderado por Philips.
- “REAAAL”: proyecto europeo.

Introducción.

- 1 Motivación.
- 2 **Estado del Arte.**
- 3 Esquema Funcional.

Proyectos a nivel Internacional

Interés en Proyectos de Investigación:

- 20 proyectos en Universidades a nivel Internacional
 - “CASAS”: Universidad Estatal de Washington.
- 15 proyectos europeos en el ámbito del FP7 y H2020.
 - AALIANCE: European ambient assisted living innovation alliance (<http://www.aalliance.eu>)

Casos de uso sobre bases de datos bien conocidas: un único usuario en un único domicilio.

Componentes Analizados en el Estado del Arte

- Sensores de Bajo Nivel (Human Activity Recognition).
- Dispositivos Móviles.
- Sensores Contextuales.
- Visión Artificial y Sonido.

Algoritmos en el Estado del Arte

- Reglas difusas: “Sistema de Inferencia Fuzzy Online Adaptativo” (AOFIS).
- Redes Neuronales para extraer patrones ADL (“Activities of Daily Living”).
- Máquinas Vector Soporte (SVM).
- Modelos Ocultos de Markov.
- Redes Bayesianas.
- Randoms Forest.
- Análisis de Series Temporales.
- Anomaly Detection (AD) / Factor Local Outlier (LOF).

Estrategia Común: Codificación semántica de eventos.

Algoritmos en la Tesis

- Reglas difusas: “Sistema de Inferencia Fuzzy Online Adaptativo” (AOFIS).
- **Redes Neuronales para extraer patrones ADL (“Activities of Daily Living).**
- **Máquinas Vector Soporte (SVM).**
- Modelos Ocultos de Markov.
- **Redes Bayesianas.**
- **Árboles de Decisión.**
- **Análisis de Series Temporales.**
- **Anomaly Detection (AD) / Factor Local Outlier (LOF).**

Arquitecturas en el Estado del Arte

- Sensórica y Redes Inalámbricas:
 - Wifi: El estándar de comunicaciones IEEE 802.11.
 - Bluetooth.
 - Zigbee estándar IEEE 802.15.4.
- Arquitectura Orientada a Servicios: Protocolos SOAP / REST.
- Tecnología OSGi.
- Anotación Semántica de la Información: RDF (Resource Description Framework).
- UniversAAL: Plataforma Europea de referencia.

Arquitecturas en la Tesis

- Sensórica y Redes Inalámbricas:
 - Wifi: El estándar de comunicaciones IEEE 802.11.
 - Bluetooth.
 - **Zigbee estándar IEEE 802.15.4.**
- **Arquitectura Orientada a Servicios: Protocolos SOAP / REST.**
- Tecnología OSGi.
- Anotación Semántica de la Información: RDF (Resource Description Framework).
- **UniversAAL: Plataforma Europea de referencia.**

Introducción.

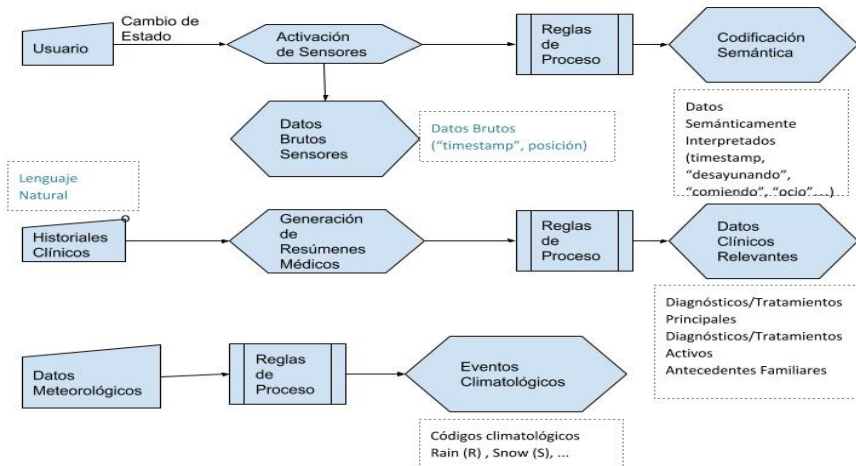
- 1 Motivación.
- 2 Estado del Arte.
- 3 **Esquema Funcional de la Propuesta.**

Módulos Funcionales de la Propuesta.

- 1 Sistema de Agregación de Datos.
 - Datos provenientes de los sensores referentes a los cambios de estado de los usuarios en un domicilio.
 - Datos obtenidos a partir de los datos clínicos de los usuarios.
 - Datos meteorológicos externos al domicilio.
- 2 Sistema Experto.
 - Módulo de Reglas Heurísticas.
 - Módulo de Detección Automática de Patrones.
- 3 Evaluación y Sistema de Notificaciones.

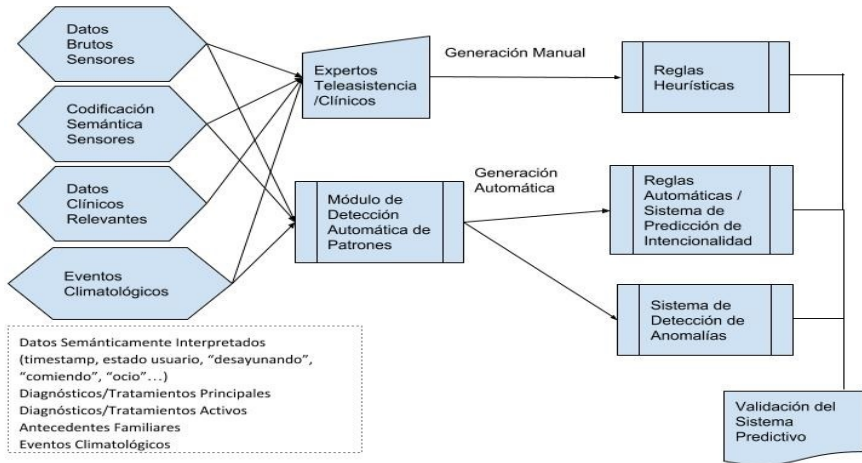
Sistema de Agregación de Datos

Sistema de Agregación de Datos



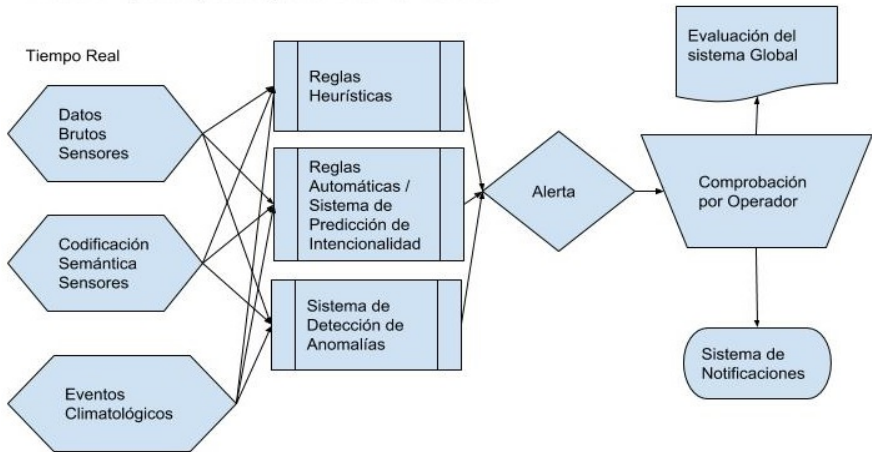
Sistemas Experto

Sistema Experto: Generación de Reglas.

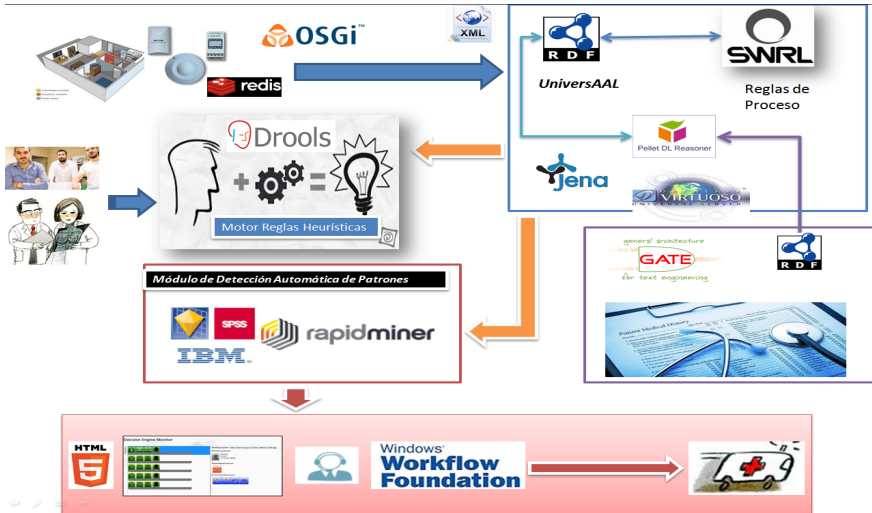


Evaluación y Sistema de Notificaciones

Sistema Experto: Aplicación y Sistema de Notificaciones.



Tecnologías Utilizadas



Contenidos

- 1 Introducción
- 2 **Generación de Resúmenes Clínicos.**
- 3 Generación de Información Sensórica.
- 4 Implementación del Sistema.
- 5 Resultados del Proyecto.
- 6 Conclusiones y Trabajos Futuros.

Generación de Resúmenes Clínicos.

- 1 Historiales Clínicos (EHR).
- 2 Algoritmo de Generación de Resúmenes.
- 3 Análisis del Corpus y la Calidad de los Datos.

Generación de Resúmenes Clínicos.

- 1 **Historiales Clínicos (EHR).**
- 2 Algoritmo de Generación de Resúmenes.
- 3 Base de los Datos y Análisis del Corpus.

Contexto de los EHR en Entornos de AAL

- Dos objetivos:
 - Contextualizar al Paciente con sus Datos Clínicos Relevantes.
 - El Contexto Clínico dirige las Recomendaciones de Control y Seguimiento de los Pacientes en el Sistema de Teleasistencia.

Componentes del Sistema

- TextMining.
- Análisis Morfosintáctico.
- Análisis Semántico.
- Ontologías y Reglas Semánticas (SWRL).
- Estándares de Codificación (UMLS).
- Interoperabilidad Semántica: CEN/ISO EN13606.

Fecha	Descripción
30/05/2016	Paciente estacionario, con mayor tendencia al sueño. Esta noche caida al intentar levantarse tras pedir lorazepam 1mg 2 ocasiones. Dolor abdominal generalizado, mal localizado. Escasa ingesta de alimentos, poco apetito, prefiere triturados. Exploración: persiste ictericia franca. Abdomen distendido, no doloroso, drenaje biliar izquierdo fuga minimamente, pero cerrado desde el sábado. En control analítico destaca BT 12,6mg/dl, leucocitosis 15000, resto de parametros similares. Explico situación clinica, sin posibilidades de otras medidas para mejora de hiperbilirubinemia, por lo que el pronostico es malo . El paciente es consciente de la situación y prefiere el manejo ambulatorio. PLAN: Contactare con Hospitalización domiciliaria de H. Zumarraga para continuar tto de paciente paliativo en situación terminal. Paciente y familia informados y conformes.
29/05/2016	Sin cambios. Pendiente de determinación de BT maana
27/05/2016	Mejoría de la ictericia. Esta noche el drenaje derecho ha tenido fuga alrededor del orificio de salida, abundante, por lo que ha precisado volver a abrir el drenaje drenado total 400ml. Dolor controlado, no nauseas ni vomitos. Refiere menor coluria en la orina y heces de aspecto normal no acolicas. Expl: abdomen globuloso, con ascitis no a tension, sin signos de peritonismo, no doloroso.

Episodio: 271517 Diagnóstico
 Fecha: 21/07/2008

Pronóstico: Márgenes quirúrgicos

Ubicación: Libro Cerrado
 Resultado: Libro

Enfermedad: Carcinoma ductal
 Ubicación: Mama derecha
 Dimensión: 4

Tratamiento: Cuadrantectomía de mama derecha
 Ubicación: Mama derecha

Enfermedad: "Carcinoma ductal in situ no infiltrado"
 Ubicación: Mama derecha

Tratamiento: RT + RH
 Componente del tratamiento: Tamoxifen 20 mg

Episodio: 271514 Terapia médica
 Fecha: 01/08/2008
Resumen(+):

Solicitud: Ecografía vaginal
 Fecha: 06/08/2008

Tratamiento: Terapia hormonal
 Componente del tratamiento: Tamoxifen 20 mg
 Resumen etapas: Inicio

Solicitud: Intervención
 Enfermedad: Hiperplasia ductal atípica
 Ubicación: Mama derecha

2013	2014	2015	2016	2017	2018	2019	2020
	05/2013-01/2015	➔	Procedimiento de exploración [Adenocarcinoma] [Hyrex n				
	05/2013-01/2015	➔	La tomografía computada de la radiografía [Adenocarcinoma]				
	05/2013-01/2015	➔	La tomografía computada de la radiografía [Neoplasma]				
	05/2013-01/2015	➔	Protocolo MAV [Neoplasias del recto]				
03/2014	➔	Procedimiento de radiología terapéutica [Adenocarcinoma] [capecitabina]					
03/2014	➔	Quimioterapia [Metastasis de la neoplasia]					
03/2014	➔	Supresión [Metastasis de la neoplasia]					
03/2014	➔	Regimen FOLFOX [Metastasis de la neoplasia] [bevacizumab]					
	05/2016	➔	Procedimiento de exploración [Carcinoma]				
	05/2016	➔	Otros procedimientos terapéuticos [Carcinoma] [Horn				

Generación de Resúmenes Médicos.

- 1 Historiales Clínicos (EHR).
- 2 ***Algoritmo de Generación de Resúmenes.***
- 3 Base de los Datos y Análisis del Corpus.

Fases del Algoritmo

- 1 Anotación y Codificación.
 - Fase 1: Extracción, Anotación y Codificación de los Conceptos Médicos.
 - Fase 2: Jerarquización de Conceptos en Grupos principales y Secundarios.
- 2 Análisis Estadístico y Semántico.
 - Fase 3: Filtrado Estadístico de los Tipos y Relaciones semánticas.
 - Fase 4: Ponderación de la Importancia Estadística.
 - Fase 5: Filtrado Semántico de la información Codificada.
- 3 Generación de Triplas.
 - Fase 6: Generación de Instancias en la Ontología.
 - Fase 7: Instanciación de los Datos Clínicos en la Ontología del Sistema.

Aportaciones en la Generación de Resúmenes Clínicos

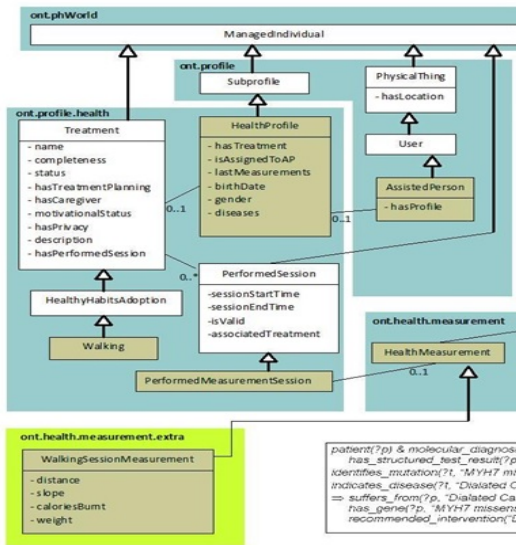
Fase 2: Jerarquización de Conceptos en Grupos Principales y Secundarios.

- Diagnósticos Principal.
- Diagnóstico Secundario.
- Tratamiento.

Fase 3: Filtrado Estadístico de los Tipos y Relaciones Semánticas.

- Asignación de Pesos Semánticos a Jerarquías de la Ontología.
- Aplicación de Algoritmos de Asociación (Apriori) sobre las Jerarquías Semánticas.

Fase 6: Generación de Instancias y Enlaces entre Ontologías.



```

@prefix ns: <http://ontology.bermatica.com/NinjaBlocks.owl#>
@prefix ns2: <http://ontology.universAAL.org/Location.owl#>
@prefix ns3: <http://ontology.universAAL.org/Device.owl#>
@prefix : <http://ontology.universAAL.org/Context.owl#> .
<[[[[[[[1]]]]]]]]> :hasProvider ns:MotionSensorContextProvider
a :ContextEvent ;
rdf:subject ns3:device55555 ;
:hasTimestamp "1415196076008"^^<http://www.w3.org/200
rdf:predicate ns3:hasValue ;
rdf:object ns3:Detected .
ns3:Detected a ns3:MotionValue .
ns3:device55555 a ns3:MotionSensor ,
ns1:Device ,
ns1:PhysicalThing ;
ns1:hasLocation [
a ns2:Room ,
ns2:HomeArea ,
ns2:Place ,
ns2:Location ;
ns2:roomFunction [
a ns2:RoomFunction
]
] ;
ns3:hasValue ns3:Detected .
:gaug a :ContextProviderType .
ns:MotionSensorContextProvider a :ContextProvider ;
:hasType :gauge .
    
```

```

patient(?p) & molecular_diagnostic_test(?t) &
has_structured_test_result(?p, ?t) &
identifies_mutation(?t, "MYH7 missense:Ser532Pro") &
indicates_disease(?t, "Dilated Cardiomyopathy")
=> suffers_from(?p, "Dilated Cardiomyopathy")
has_gene(?p, "MYH7 missense:Ser532Pro")
recommended_intervention("DCM Monitoring and Management")
    
```

Generación de Resúmenes Médicos

- 1 Historiales Clínicos (EHR)
- 2 Algoritmo de Generación de Resúmenes.
- 3 *Base de los Datos y Análisis del Corpus.*

Base del Análisis del Corpus

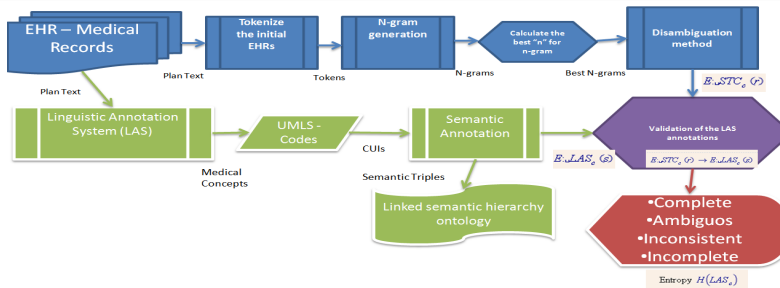
Uso del sistema de Generación Automática de Resúmenes Clínicos:

- La Triangulación y Segmentación de pacientes, de cara a mejorar el Análisis Diagnóstico.
- Análisis y recomendación de tratamientos en la mejora de la Eficiencia Terapéutica .
- Análisis y mejora de Procesos Asistenciales.

Análisis del Corpus

- No existe un corpus médico de referencia para validar las anotaciones semánticas.
- Generamos un anotador estadístico (STC) que comparamos con el anotador semántico (LAS).
- Existe una fase de desambiguación de n-gramas.
- La comparativa entre ambos codificadores se mide en función de la entropía de Shannon, según las siguientes leyes:
 - (i) El número promedio de conceptos de un sistema de anotación debe ser mayor o igual que la entropía H de la fuente.
 - (ii) Si el número de conceptos es suficientemente grande, existe un sistema de anotación único para la fuente, de modo que el número promedio de conceptos del sistema de anotación sea tan cercano a H como se desee.

Análisis del Corpus



- $STC_c = \{A_1, A_2, \dots, A_i, \dots, A_n\}$ es un vector de n conceptos anotados estadísticamente.
- $LAS_c = \{B_1, B_2, \dots, B_j, \dots, B_m\}$ es un vector de n conceptos anotados semánticamente (UMLS).
- $E : STC_c(r) \rightarrow E : LAS_c(s)$, codifica bloques de r conceptos de STC_c como s conceptos de LAS_c ,
- La entropía de LAS_c viene dada por $H(LAS_c) = \sum_{i=1}^s f_{B_i} \log \frac{1}{f_{B_i}}$, donde f_{B_i} es la frecuencia del concepto B_i en el vector de LAS_c

Tabla de Validación entre Anotadores

Output Example $E : STC_c(r) \rightarrow E : LAS_c(s)$	Entropy $H(LAS_c)$	Type
B1 {Right,breast,cancer} $s = 1$	$1 \cdot \log_2 1 = 0$ There is not entropy	Complete ($H = 0$)
B1 (Right,breast), B2(\emptyset) $s = 2$	$(\frac{1}{2} \log_2 2) = 0.5$	Ambiguous ($0 < H < 1$)
B1 (Right), B2(\emptyset), B3(\emptyset) $s = 3$	$(\frac{1}{3} \log_2 3) + 0 + 0 = 0.52$	Ambiguous ($0 < H < 1$)
B1 (Right), B2 (cancer, breast) $s = 2$	$(\frac{1}{2} \log_2 2) + (\frac{1}{2} \log_2 2) = 1$	Inconsistent ($H > 1$)
B1(Right), B2(breast), B3 (cancer) $s = 3$	$\frac{1}{3} \log_2 3 + \frac{1}{3} \log_2 3 + \frac{1}{3} \log_2 3 =$	Inconsistent ($H > 1$)

Contenidos

- 1 Introducción.
- 2 Generación de Resúmenes Clínicos.
- 3 **Generación de Información Sensórica.**
- 4 Implementación del Sistema.
- 5 Resultados del Proyecto.
- 6 Conclusiones y Trabajos Futuros.

Generación de Información Sensórica.

- 1 Tipología de Sensores.
- 2 Preprocesado de los Datos.

Generación de Información Sensórica.

- 1 **Tipología de Sensores.**
- 2 Preprocesado de los Datos.

Sensores Integrados en el Sistema

- Sensores corporales: frecuencia cardíaca, frecuencia respiratoria, temperatura de la piel, la posición / postura
- Sensores ambientales simples
 - Sensores de Presencia PIR
 - Sensores de Puertas
 - Sensores de Humo/Gas
- Sensores ambientales avanzados
 - Sensor de Video-Análisis
 - Micrófono de Detección de ruido ambiental
 - Sensor de ocupación de cama

Pruebas preliminares en laboratorio del Sistema Experto

- Exigencia en la verificación técnica y las pruebas preliminares íntegras en un entorno de laboratorio.
- Las pruebas de laboratorio tuvieron lugar en un “Living Lab” o entorno controlado.
- Supervisadas por un equipo auditor externo (H-ENEA).
- Perfil de usuarios: jubilados, no dependientes, con chequeos médicos regularmente y necesidad de acompañamiento.

Conclusiones: Sólo sensores ambientales simples.

Generación de Información Sensórica.

- 1 Tipología de Sensores.
- 2 **Preprocesado de los Datos.**

Preproceso de los Datos

- Monitorizados 60 domicilios.
 - En cada hogar habrá una o varias personas viviendo habitualmente.
 - Los usuarios pueden tener varias visitas, ya sea cuidadores o miembros de la familia.
 - Sólo los cambios de estado de los sensores o la medición se guardan en la base de datos de análisis
 - Si no hay cambios, se registra el mismo estado o medición en la base de datos cada 5 minutos
 - En total, se han almacenado 556.972 eventos.
- Dos tipos de diseños experimentales:
 - Análisis con datos semánticamente codificados.
 - Análisis con datos no codificados.

- 1 Introducción
- 2 Generación de Resúmenes Clínicos
- 3 Generación de Información Sensórica
- 4 Implementación del Sistema**
- 5 Resultados del Proyecto
- 6 Conclusiones

Implementación del Sistema.

- 1 Sistema Local.
- 2 Módulo de Reglas.
- 3 Módulo de Detección Automática de Patrones.
- 4 Metodología de Control y Evaluación.

Implementación del Sistema.

- 1 **Sistema Local.**
- 2 Módulo de Reglas.
- 3 Módulo de Detección Automática de Patrones.
- 4 Metodología de Control y Evaluación.

Sistema Local

- Recopilar y agregar los datos de todos los sensores en una base de datos local no-sql.
- Chequear que los sensores está activos.
- Intercambiar datos entre el Sistema Local y el Sistema en “Cloud”.
- Raspberry Pi + UniversAAL + OSGi + Redis.



Implementación del Sistema.

- 1 Sistema Local.
- 2 **Módulo de Reglas.**
- 3 Módulo de Detección Automática de Patrones.
- 4 Metodología de Control y Evaluación.

Módulo de Reglas

- Módulo de Reglas de Proceso: Inferencia Semántica (SWRL).
- Módulo de Reglas Heurísticas: Algoritmo de Rete (Drools).

Módulo de Reglas de Proceso

- Gestionan la Calidad de los Datos y la creación de Eventos (SWRL).
 - Codificación Semántica de los Valores de los Sensores.
 - Generación de los Resúmenes Médicos.
 - Integración entre los Sistemas.
- Se infieren directamente en la Ontología del Sistema.

Edit [X]

Name
§1

Comment
§1 esta en la habitación de noche está durmiendo

Status

Ok

```
ns1:Sensor(?x) ^ autogen0:roomFunction(?x, autogen0:SleepingRoom) ^ ns1:hasTime(?x, ?hora) ^ swrlb:greaterThan(?hora, 0) ^ swrlb:lessThan(?hora, 7) -> ns1:hasSemanticValue(?x, 'S')
```

Cancel Ok

Property assertions: Evento_234 [11:48:02]

Object property assertions +

- roomFunction SleepingRoom
- roomFunction SleepingRoom

Data property assertions +

- hasTime "1"^^xsd:int
- hasSemanticValue "S"^^xsd:string
- hasTime "1"^^xsd:int

Explanation 1 Display laconic explanation

Explanation for: Evento_234 hasSemanticValue "S"^^xsd:string

- 1) Evento_234 hasTime "1"^^xsd:int In ALL other justific.
- 2) Evento_234 roomFunction SleepingRoom In ALL other justific.
- 3) roomFunction Domain Sensor In NO other justific.
- 4) Sensor(?owl:x, roomFunction(?owl:x, SleepingRoom), hasTime(?owl:x, ?owl:hora), swrlb:greaterThan(?owl:hora, 0), swrlb:lessThan(?owl:hora, 7), hasSemanticValue(?owl:x, "S"^^xsd:string) In ALL other justific.

Módulo de Reglas Heurísticas

- Conocimiento de expertos (Drools).

```
rule 'Caida_Night_Levanta_Cama'
when
    ed : EventDetection ( DiaSem : sDayOfWeek, numDate : EventDate, numTime : EventTime, numTimeTicks : EventDateT
    con : ConfigTiempos ( ConfType == "NoActivityNight", MinConf : ConfTime )
    o3 : Evento ( SensorTypeName == "EnvDoorContact", HoraPuerta : MinutesFirstEvent )
    o2 : Evento ( SensorTypeName == "EnvPresion", SensorValue == 0 )
    o : Evento ( SensorTypeName == "EnvPresence", eventRoomType == "Dormitorio", EventRoomPrincipal == "1", MinutesLa
    not AlarmExceptions ( StartDate <= numDate, EndDate >= numDate)
    not AlarmExceptions ( sDayOfWeek == DiaSem)
    a : AlarmExceptions ( StartTimeTicks <= numTimeTicks, EndTimeTicks >= numTimeTicks )

then
    Controller.NotificarAlarma(24, o.endEventDateTime, o.MonitoringPersonId, 8);
    Console.WriteLine("Caida_Night_Levanta_Cama");
end

rule 'Caida_Night_Bathroom'
when
    ed : EventDetection ( DiaSem : sDayOfWeek, numDate : EventDate, numTime : EventTime, numTimeTicks : EventDateT
    con : ConfigTiempos ( ConfType == "LongActivityNight", MinConf : ConfTime )
    o3 : Evento ( SensorTypeName == "EnvDoorContact", HoraPuerta : MinutesFirstEvent )
    o2 : Evento ( SensorTypeName == "EnvPresion", SensorValue == 0 )
    o : Evento ( SensorTypeName == "EnvPresence", eventRoomType == "Baño", MinutesLastEvent < (HoraPuerta-5), Minutes
    not AlarmExceptions ( StartDate <= numDate, EndDate >= numDate)
    not AlarmExceptions ( sDayOfWeek == DiaSem)
    a : AlarmExceptions ( StartTimeTicks <= numTimeTicks, EndTimeTicks >= numTimeTicks )

then
    Controller.NotificarAlarma(24, o.endEventDateTime, o.MonitoringPersonId, 9);
    Console.WriteLine("Caida_Night_Bathroom");
end
```


Implementación del Sistema.

- 1 Sistema Local.
- 2 Módulo de Reglas.
- 3 **Módulo de Detección Automática de Patrones.**
- 4 Metodología de Control y Evaluación.

Módulo de Detección Automática de Patrones

- 1 Análisis con Datos Semánticamente Codificados.
- 2 Análisis con Datos No Codificados.

Preprocesado de Datos Semánticamente Codificados

- Proceso de “codificación” de los datos de los sensores en etiquetas.
- Los datos en bruto de los sensores son procesados por el sistema experto para determinar cuál es el contexto del anciano.
 - Dormir (S), Cocinar (C), Comer (E), Hacer el trabajo doméstico (D), salir al aire libre (O), Deporte al aire libre (U), Usar la Tablet (T), Tiempo de Ocio (P), Hablar por el móvil (X).
 - Agregación datos externos climatológicos: Haze (C), Fog (N), Low Fog (N), Fog (I), Precipitation (P), Drizzle (L), Rain (U), Torn Rain (V), Tornado Sight (R), Rain Shower (H), Rain (E), Snow (E), Shower Hail (T), Freezing Rain (T).
- Las fechas de los datos históricos recogidos por el sistema se formalizan en nuevos campos tales como la semana, el día del mes, la hora, el cuarto horario de la acción, el trimestre, y el día de la semana.

Sistema de Clasificadores

Se aplica un modelo híbrido (Método de Votación Ponderada):

- Árboles de Decisión (C5.1, Quest, C&R, CHAID)
- Red Neuronal
- Red Bayesiana
- Regresión Logística
- SVM

Selección Paramétrica: Algoritmo Genético (generaciones = 50, individuos = 5, tipo de mutación: gaussiana, tipo de selección: por torneo, ratio de crossover = 0.25).

- 1 **Análisis con Datos Semánticamente Codificados.**
- 2 **Análisis con Datos No Codificados.**

Preproceso de Datos en Bruto

- Inclusión de un nuevo Indicador: Cambio_de_Ubicación.
- Inclusión de Frecuencias de Duración en el Evento Actual, en el Evento Anterior.
- Para cada Domicilio, se genera una Tabla de Discretización de Frecuencias por el Método “Intervalos de Cuantil”.

Quantile Grouping Method

Home: 31			Home: 43		
Interval	Below	Upper	Interval	Below	Upper
1	>= 15,01666667	< 18,83333333	1	>= 15,03333333	< 17,76666667
2	>= 18,83333333	< 23,9	2	>= 17,76666667	< 20,8
3	>= 23,9	< 31,41666667	3	>= 20,8	< 24,76666667
4	>= 31,41666667	< 39,33333333	4	>= 24,76666667	< 29,81666667
5	>= 39,33333333	< 49,83333333	5	>= 29,81666667	< 36,21666667
6	>= 49,83333333	< 62,1	6	>= 36,21666667	< 44,63333333
7	>= 62,1	< 79,6	7	>= 44,63333333	< 59,05
8	>= 79,6	< 106,81666667	8	>= 59,05	< 90,75
9	>= 106,81666667	< 175,58333333	9	>= 90,75	< 177,7

Clasificador de Cambio de Ubicación

Cambio_Ubicación_Usuario = modelo_binario(Ubicación_previa, frecuencia_ubicación_anterior, frecuencia_ubicación_actual, día, hora, semana_mes)

Resultados para el campo de resultado Cambio

Comparando \$XS-Cambio con Cambio

Correcto	1.217	80,49%
Erróneo	295	19,51%
Total	1.512	

Matriz de coincidencias para \$XS-Cambio (las filas muestran las reales)

	0	1
0	477	115
1	180	740

Resultados para el campo de resultado Cambio

Comparando \$XS-Cambio con Cambio

Correcto	1.226	75,82%
Erróneo	391	24,18%
Total	1.617	

Matriz de coincidencias para \$XS-Cambio (las filas muestran las reales)

	0	1
0	28	365
1	26	1.198

Dos tipos de alertas:

- Usuario no está en la misma Ubicación esperada (Tipo I).
- No hay Cambio de Ubicación cuándo se predice (Tipo II).

Clasificador de Ubicación del Usuario

$Ubicación_Usuario = \text{modelo}(Ubicación_previa, frecuencia_Ubicación_previa, frecuencia_Ubicación_Actual, día, hora, semana_mes)$

Modelo	min.	Exactitud
CHAID	5	65%
Red Bayesiana	5	81,413%

- Confianza conjunta de ambos clasificadores (Vote).
- “k-fold” suministra el “**Umbral de Confianza Personalizado**” por Domicilio.

Clasificador Datos No Codificados

Resultados para el campo de resultado Cambio

Comparando \$XS-Cambio con Cambio

Correctos	198	87,22%
Erróneos	29	12,78%
Total	227	

•Predicción(Cambio_Ubicación_Usuario) ∈ {0,1}

Evaluación del rendimiento

Informe de valores de confianza para \$XSC-Cambio

Resultados para el campo de resultado Estado_Usuario

Comparando \$XS-Estado_Usuario con Estado_Usuario

Correctos	195	85,9%
Erróneos	32	14,1%
Total	227	

•Predicción(Ubicación_Usuario)

Evaluación del rendimiento

Baño 1	3,451
Entrada	2,47
Habitacion 1	1,204
Habitacion 2	1,25
Puerta	1,975
Salon	1,13

Informe de valores de confianza para \$XSC-Estado_Usuario

Rango	0,194 - 0,853
Media para correctos	0,538
Media para incorrectos	0,374
Siempre correctos por encima de	0,8 (3,96% de casos)
Siempre incorrectos por debajo de	0,218 (1,32% de casos)
90,1% Precisión por encima de	0,327
2,0 Veces correctas por encima de	0,403 (93,08% de casos)

Umbral de Confianza Personalizado
por cada Domicilio

Funcionamiento del Sistema “Cloud”

Paso 1: Preproceso de los Datos Brutos por Domicilio.

- Paso 2: Ejecución las Reglas Heurísticas para detectar Alertas Deterministas.

Paso 3: Iniciar el Proceso de Aprendizaje Automático, Diario, en base a los Históricos.

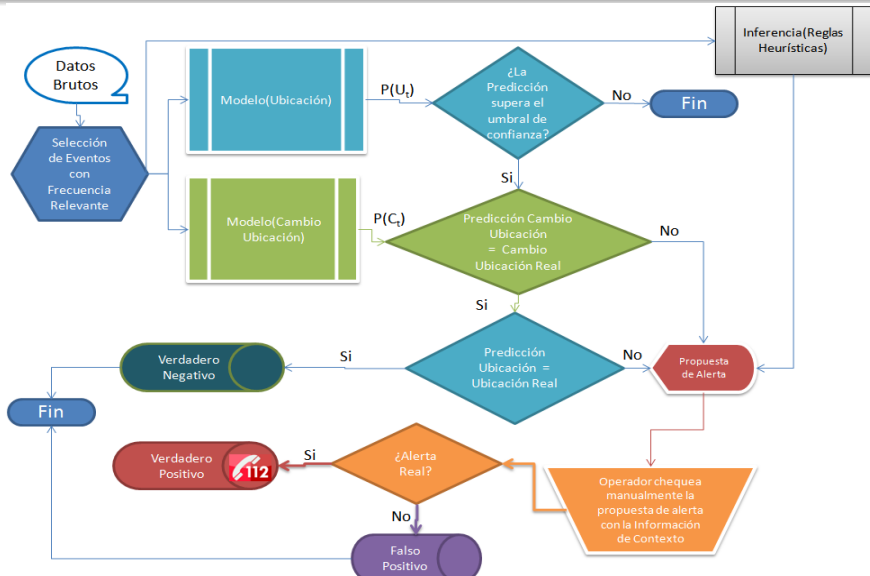
- Paso 4: Aplicar el Modelo de Comportamiento Aprendido en el paso previo, sobre los datos del Paso 1: Soporte y Confianza.
- Paso 5: Comparar las Predicciones con las Ubicaciones Reales, y determinar si existe Probabilidad de Alerta.

Paso 6: En el caso de Probabilidad de Alerta, Verificarla Manualmente, y almacenar los resultados en el Histórico de Eventos, (VP, FP, VN).

Paso 7, Comenzar desde el Paso 1.

Implementación del Sistema.

- 1 Sistema Local.
- 2 Módulo de Reglas.
- 3 Módulo de Detección Automática de Patrones.
- 4 **Metodología de Control y Evaluación.**



- 1 Introducción.
- 2 Generación de Resúmenes Clínicos.
- 3 Generación de Información Sensórica.
- 4 Implementación del Sistema.
- 5 **Resultados de la Tesis.**
- 6 Conclusiones y Trabajos Futuros.

Resultados de la Generación de los Resúmenes Clínicos

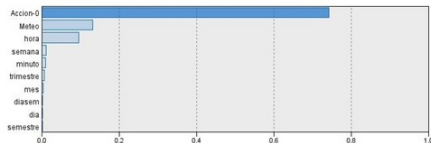
- 20.000 términos en dos contextos diferentes.
- 4.500 conceptos codificados/contexto.

Annotation keys	Breast Cancer Context	Colorectal Cancer Context
Complete	74.9%	82.84%
Inconsistent	17.5%	10%
Ambiguous	5.9%	5.6%
Incomplete (Fault)	1.7%	1.56%

- Precisión: 92.62%.
- Sensibilidad: 78.87%

Resultados del Clasificador de Datos Semánticamente Codificados

$Estado_Usuario_n = \text{modelo}(\text{modelomes}, \text{semana_mes}, \text{dia_semana}, \text{hora}, \text{cuarto_horario}, \text{trimestre}, \text{estado_meteorológico}, \text{Estado_Usuario}_{n-1})$



accuracy: 81.80%

	true S	true P	true E	true T	true O	true C	true D	true U	true X	precision	class
pred. S	7404	270	9	2	74	23	0	0	0	3 95.11%	
pred. P	313	2318	136	51	66	207	23	2	0	0 74.39%	
pred. E	16	175	851	84	6	68	124	0	0	0 64.27%	
pred. T	0	73	17	994	126	27	6	38	1	1 77.54%	
pred. O	0	87	44	6	843	119	1	72	14	14 71.08%	
pred. C	36	197	244	42	3	1321	56	0	0	0 69.56%	
pred. D	2	25	0	83	8	75	345	0	0	0 64.13%	
pred. U	0	28	2	0	59	33	2	242	5	5 65.23%	
pred. X	0	0	0	0	1	0	0	3	2	2 33.33%	

Resultados del Clasificador de Datos No Codificados

- Exactitud Media del Clasificador de Cambio de Estado: 83,87%.
- Exactitud Media del Clasificador de Ubicaciones: 63%.
- Exactitud Media: 73,44%.

Aquellas viviendas con una exactitud cuya confianza está por encima del 75%, necesitan una media de 28 días de histórico para el aprendizaje.

Comparativa Codificación Semántica vs Sin Codificación

Sistema	Exactitud Media
Análisis con Datos Codificados Semánticamente	81.80%
Análisis con Datos No Codificados Semánticamente	73,44%

Comparativa Sistema de Reglas vs Módulo de Detección Automática de Patrones

Sistema	Exactitud	Precisión
Módulo de Reglas Heurísticas	74%	12%
Módulo de Detección Automática de Patrones	98%	59%

Publicaciones (1)

JCR.

- “Lynx: Automatic Elderly Behavior Prediction in Home Telecare”. BioMed Research International, vol. 2015, Article ID 201939, 18 pages, 2015. doi:10.1155/2015/201939.
- “A novel methodology for clinical semantic annotations assessment”. Journal of Applied Logic. Elsevier. 2018. "In Press".

Otras.

- “An Automatic Telemonitoring System for Elderly People at Home”. International Journal of Sensors Wireless Communications and Control (SWCC) Volumen: 4 Número: 2 Páginas: 57-66.

Congresos

- “Real implantation of an expert system for elderly home care”. 10th International Conference on Hybrid Artificial Intelligence Systems (HAIS 2015). 22th-24th June, 2015.
- “Real prediction of elder people abnormal situations at home”. 11th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2016). 19th – 21st October, 2016.
- “Behavior prediction in home telecare systems.” Ponencia invitada. 6th International Conference on Applied Informatics and Computing Theory (AICT 2015). Salerno, Italy.

- 1 Introducción.
- 2 Generación de Resúmenes Clínicos.
- 3 Generación de Información Sensórica.
- 4 Implementación del Sistema.
- 5 Resultados de la Tesis.
- 6 **Conclusiones y Trabajos Futuros**

Conclusiones (I)

- En este trabajo está enfocado en la automatización de los sistemas asistenciales.
- La seguridad de los ancianos en los domicilios pasa por un sistema automático de análisis de patrones.
- El sistema soporta la integración de información relevante clínica, extraída directamente de texto no estructurado.
- El sistema permite la inclusión de Reglas de Expertos para el control de la actividad.

Conclusiones (II)

- El sistema complementa el conocimiento experto en base a un Sistema de Predicción de Intencionalidad desarrollado con técnicas de Inteligencia Artificial.
- Se utiliza una medida de control “Umbral de Confianza” para evitar “FP”.
- El sistema necesita una media de 28 días de aprendizaje.
- Se demuestra que el Sistema Automático mejora la precisión y exactitud en la gestión de alertas del Sistema de Reglas.

Trabajos Futuros

- Expandir el sistema a más domicilios.
- Reforzar los sistemas inteligentes para mejorar la exactitud y precisión del sistema.
- Tratar de generalizar el conocimiento de algunos domicilios a otros en sus fases iniciales.
- Transformar esta propuesta en una referencia en las plataformas de teleasistencia en casa, reales y comercializables.
- Permitir a la comunidad investigadora y social de una herramienta que permita mejorar el conocimiento en sistema AAL.

¡Muchas gracias por su atención!

Corollary

Proyectos a nivel Internacional

- “CASAS”: Universidad Estatal de Washington. “
- “Aging in Place”: Universidad de Missouri.
- “Aware Home project”: Georgia Institute of Technology.
- “DOMUS”: Universidad de Sherbrooke.
- “House_n”: Massachusetts Institute of Technology.

Proyectos a nivel Europeo

- “CONFIDENCE (FP7 ICT 7.1)”: Cuidados y atención para las personas detectando situaciones anormales.
- “PERFORM (FP7 ICT 5.1)”: Monitorización de la evolución de enfermedades mediante el empleo de sensores.
- “HERMES”: Cognitive care and guidance for active aging.
- “AALIANCE”: European ambient assisted living innovation alliance.
- “MonAMI”: Mainstreaming on Ambient Intelligence.
- “SOPRANO”: Desarrollo de entornos inteligentes dirigidos a las personas mayores.
- “PERSONA”: Espacios perceptivos que promuevan el envejecimiento independiente.
- “NETCARITY”: Atención sanitaria y seguridad en el domicilio.

Fase 1: Extracción, Anotación y Codificación de los Conceptos Médicos

- “Tokenizer” por frases/palabras (GATE).
- Normalización de Conceptos:
 - Diccionario “stopwords”.
 - Análisis morfosintáctico (POS).
 - Lematización (Tagger).
 - Diccionarios Médicos, Acrónimos (Javier Yetano), Personalizados, Equivalencias (Snomed-CT).
 - Corrección ortográfica (Levenshtein).
 - Generación de n-gramas.
- Codificación UMLS (MetaMap).
- Análisis del Corpus para la corrección de errores.

Fase 2: Jerarquización de conceptos en grupos principales y secundarios

- Cada uno de estos códigos UMLS se asigna a un grupo principal en la ontología del trabajo:
 - Diagnósticos Principal.
 - Diagnóstico Secundario.
 - Tratamiento.
- Esta asignación requiere:
 - un mecanismo adecuado de selección (Reglas Semánticas SWRL).
 - un motor de inferencia (Pellet)..
- Se utilizan dos fuentes externas de anotación:
 - las jerarquías de codificación UMLS.
 - un diccionario médico de sinónimos y acrónimos.

Fase 3: Filtrado estadístico de los tipos semánticos y relaciones semánticas

- Los conceptos codificados sigue una distribución denominada “distribución Zipf”.
- $P(x) = \frac{x^{-(p+1)}}{\zeta(p+1)}$, $\zeta(z)$ es la función zeta de Riemann, y p es el parámetro que mide la velocidad de decaimiento de la distribución Zipf.
- El test Z-value $Z_i = \frac{|X_i - \mu|}{\sigma}$ provee una buena medida de detección de outliers en una distribución Zipf ($Z_i \geq 3$).
- Se seleccionan los tipos semánticos resultantes de la selección.
- Se analizan las relaciones entre tipos semánticos: Algoritmo de Asociación (Apriori).

Fase 4: Ponderación de la importancia estadística

- Se vectoriza cada episodio en función de la distribución Tf/idf.
- A cada concepto se le asigna un peso en función de su peso relativo en el vector tf/idf y su tipo semántico.
- A cada episodio se le asigna el peso como suma ponderada de sus componentes.
- El peso de cada episodio sirve para generar los resúmenes clínicos.

Fase 5: Filtrado semántico de la información anotada

- Generación relaciones semánticas en función de las asociaciones Fase 3.
- Asignación a cada relación de una propiedad en la Ontología Clínica.
- Inclusión de asignaciones en forma de Reglas de Proceso semánticas (SWRL).

Fase 6: Generación de instancias en la ontología

- Generación de grafos semánticos:
 - Cada par de conceptos se transforman en una unión de relación (un grafo).
 - Cada grafo se corresponde a una categoría diferente de información (por ejemplo, enfermedades, síntomas y signos o medicamentos).
 - Ejecución de las Reglas de Proceso por un razonador: Pellet.

Fase 7: Instanciación de los Datos Clínicos en la Ontología del Sistema

- Los resúmenes automáticos generados en la Fase 6 alimentan a la Ontología de Telesistencia (UniversAAL).
- Propiedad “same_as” del estándar W3 semántico.
- Acceso desde consultas SPARQL.

Algoritmo Rete (I)

- Creado por el Dr. Charles L. Forgylf.
- Ampliamente utilizado en Sistemas Expertos (CLIPS, Drools, JESS, ...)
- Se trata de un método eficiente para comparar una gran colección de patrones con una gran colección de objetos.
- La red que genera Rete es un grafo dirigido sin ciclos conformado por nodos que representan las condiciones de las reglas.
- La red se compone de dos subredes: la red Alfa y la red Beta.
 - La red Alfa se compone de nodos de una sola entrada, que representan los elementos de una condición.
 - La red Beta, formada por nodos de dos entradas, define las uniones entre elementos.

Algoritmo Rete (II)

- Ventajas
 - Elimina la duplicación en la Reglas.
 - No aplica dos veces la misma Regla a los mismos argumentos.
 - Preferencia a Reglas más específicas.
 - Permite Prioridad de Operación.
- Desventajas: Consume gran cantidad de memoria para incrementar velocidad de procesamiento.

Inferencia Semántica. Ontologías

- Ontología: Representación semántica del conocimiento.
- Se define como una colección de todas las estructuras que respetan un conjunto de restricciones impuestas en un contexto.
- Una ontología corresponde a un conjunto de fórmulas de la Lógica de Primer Orden.

$X \in \text{Axiom}$::=	$\langle \text{uri} \rangle R$	
$R \in \text{Rule}$::=	$A_1 \wedge \dots \wedge A_m \Rightarrow B_1 \wedge \dots \wedge B_n,$	$m, n \geq 0$
$A, B \in \text{Atom}$::=	$C(i)$	(Descripción)
		$C(d)$	(Datos)
		$P(i_1, i_2)$	(Prop. objeto)
		$P(i, d)$	(Prop. datatype)
		$\text{sameAs}(i_1, i_2)$	(Igualdad)
		$\text{differentFrom}(i_1, i_2)$	(Desigualdad)
		$\text{builtIn}(r, d_1, \dots, d_k)$	$k \geq 0$ (Función)
$i \in \text{Individuo}$::=	$v_i \mid \text{id}$	
$d \in \text{Datatype}$::=	$v_d \mid \text{literal}$	

Inferencia Semántica. SWRL (I)

- SWRL: Razonamiento en la Web Semántica: Semantic Web Rule language (SWRL)
- SWRL extiende las ontologías (OWL) con reglas (RuleML)
 - Define axiomas lógicos que deben ser ciertos
 - Los axiomas deben estar formados por una URI y una regla
 - Las inferencias se limitan a propiedades estructurales
 - Todas las reglas R son de la forma antecedente \Rightarrow consecuente formados por conjunciones $a_1 \wedge a_2 \wedge \dots \wedge a_i$ ($i \geq 0$)
 - Los individuos (objetos) y datatypes (valores) se tratan por separado

Inferencia Semántica. SWRL (II)

- Ventajas
 - Incorpora las inferencias directamente en la Ontología.
 - Las reglas son aplicables a cualquier nivel de la Ontología.
 - Se permiten reglas sobre objetos y propiedades
- Desventajas:
 - No permiten incorporar probabilidad a las reglas.
 - No permiten Prioridades.

Sistemas de Aprendizaje Automático

Existen MDSS no basados en el conocimiento experto, que aprenden de los datos brutos (semi / no estructurados), y se basan en técnicas probabilísticas: los patrones se toman como ejemplos o casos en el pasado y el sistema tiene capacidad de aprendizaje probabilístico.

- Técnicas
 - Algoritmos Supervisados
 - Algoritmos No Supervisados.
- Funcionalidad
 - Descriptiva
 - Predictiva
 - Prescriptiva
 - Proactiva

Árboles de Decisión (I)

En un árbol de decisión, cada nodo del árbol es un atributo y de él parten o nacen tantas ramas como valores puede tener ese atributo.

- Utilidades
 - Los ejemplos del juego de datos pueden ser representados como pares de 'valor-atributo'.
 - La función objetivo toma valores discretos.
 - Las hipótesis pueden ser expresadas mediante disyunciones.
- Funcionamiento: Elección del mejor criterio para la división de los datos en base a la Entropía.

$$Entropia(S) = \sum_{(i=1)}^n - p_i \log_2 p_i$$

$$Ganacia(S, A) \equiv Entropia(S) - \sum_{v \in \text{Valores}(A)} \frac{|S_v|}{|S|} Entropia(S_v)$$

Árboles de Decisión (II)

Ventajas:

- Las reglas que generan son fáciles de interpretar.
- Escalan mejor que otros tipos de técnicas.
- Tratan bien los datos con ruido.

Inconvenientes:

- No manejan de forma sencilla los atributos continuos.
- Tratan de dividir el dominio de los atributos en regiones rectangulares y no todos los problemas son de ese tipo.
- Tienen dificultad para trabajar con valores perdidos.
- Pueden tener problemas de sobreaprendizaje.
- No detectan correlaciones entre atributos.

Redes Neuronales (I)

- Modelo computacional basado en un gran conjunto de unidades neuronales simples (neuronas artificiales).
- Las neuronas se enlazan entre capas adyacentes mediante un conjunto de conexiones.

$$z_j = h(a_j) = h\left(\sum_{i=1}^n w_{ji}x_i + b_j\right)$$

- Backpropagation

$$ep^2 = \frac{1}{2} \sum_{k=1}^l (\delta_k^2)$$

ep^2 : error cuadrático medio para cada patrón de entrada

δ_k^2 : error en la neurona k de la capa de salida con l neuronas.

Redes Neuronales (II)

Ventajas

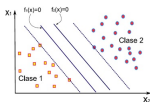
- Auto organización.
- Tolerancia a fallos.
- Flexibilidad.
- Tiempo real.

Inconvenientes:

- Complejidad de aprendizaje para grandes tareas.
- No permite interpretar lo que se ha aprendido.
- Los sobreentrenamientos deben ser controlados en la fase de aprendizaje.

Máquinas de Vector Soporte (SVM) (I)

La máquina de vectores soporte es un método de aprendizaje basado en muestras para la realización de clasificadores y regresores. Generaliza el método “generalized portrait”, propuesto por Vapnik y Lerner (Vapnik y Lerner, 1963) para la resolución de problemas de clasificación linealmente separables mediante lo que se denomina hiperplano óptimo de separación.



El objetivo es encontrar una función lineal $f(x) = (w, x) + b$, entre todos los hiperplanos canónicos que clasifican correctamente los datos, aquel con menor norma, o, equivalente, como mínimo

$$\text{Min}_{w,b} \frac{1}{2} \|w\|^2$$

Máquinas de Vector Soporte (SVM) (II)

Ventajas

- Una excelente capacidad de generalización, debido a la minimización del riesgo estructurado.
- Existen pocos parámetros a ajustar.
- Evita la existencia de un mínimo local.
- El modelo final puede ser escrito como una combinación de vectores de soporte.

Inconvenientes:

- La predicción del clasificador no tiene significado probabilístico.
- Problemas computacionales al aplicarse sobre conjuntos grandes de datos de entrenamientos.

Redes Bayesianas (I)

Basadas en el teorema de Bayes: expresa la probabilidad condicional de un evento alatorio.

Sea $\{A_1, A_2, \dots, A_n\}$ un conjunto de sucesos mutuamente excluyentes y exhaustivos, tales que la probabilidad de cada uno de ellos es distinta de cero. Sea B un suceso cualquiera del que se conocen las probabilidades condicionadas $P(A_i|B_i)$. Entonces, la probabilidad $P(A_i|B)$ viene dada por la expresión:

$$P(A_i|B) = \frac{P(B|A_i) \cdot P(A_i)}{P(B)} \quad P(B) = \sum_{i=1}^n P(B|A_i) \cdot P(A_i)$$

$$P(A_i|B) = \frac{P(B|A_i) \cdot P(A_i)}{\sum_{i=1}^n P(B|A_i) \cdot P(A_i)}$$

Redes Bayesianas (II)

Ventajas

- Admiten el aprendizaje sobre relaciones de dependencia y causalidad.
- Permiten la combinación de conocimiento con datos.
- Evitan el sobreajuste continuo de los datos y pueden manejar bases de datos incompletas.
- Permiten realizar inferencias bidireccionales.

Desventajas

- Asumir que las variables tienen independencia condicional respecto a la clase conlleva a una falta de precisión.

La Validación Cruzada

Existen diversas técnicas para validar los métodos de clasificación, como son:

- La comparación de los resultados mediante modelos físicos teóricos o con simulaciones.
- La utilización de nuevos conjuntos de datos conocidos para comparar con los obtenidos.
- El uso de técnicas de validación cruzada.
 - hold-out.
 - k-fold (k veces hold-out).

- Precisión = $\frac{tp}{(tp+fp)}$

- Recall (Sensibilidad) = $\frac{tp}{(tp+fn)}$

- Exactitud (Accuracy) = $\frac{tp+tn}{(tp+tn+fp+fn)}$

Detección de Anomalías: Local Outlier Factor

Algoritmo no supervisado.

Busca datos atípicos de forma local.

Proporciona un valor de cuánto de atípico es cada punto en el espacio vectorial.

Basado en la diferencia de densidad (lrd) que tiene con respecto a sus vecinos más cercanos.

$$lrd_{MinPts}(p) = \left(\frac{\sum_{o \in N_{MinPts}(p)} reach-dist_{MinPts}(p, o)}{|N_{MinPts}(p)|} \right)^{-1}$$

$$LOF_{MinPts}(p) = \frac{\sum_{o \in N_{MinPts}(p)} \frac{lrd_{MinPts}(o)}{lrd_{MinPts}(p)}}{|N_{MinPts}(p)|}$$

Series Temporales: Modelo ARIMA

- Detección de patrones de repetición en secuencia temporales
- Descomposición de la variación de una serie en varias componentes básicas $X_t = T_t + E_t + I_t$
- ARIMA (Modelos Autorregresivos Integrados de Medias Móviles)

- Modelos Autorregresivos $Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t$

- Modelos de Medias Móviles

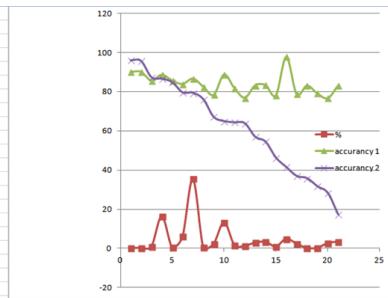
$$Y_t = \mu + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q}$$

- Modelos ARIMA (p,0,q)

- Modelos ARIMA(p, d, q)

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

Home	%	Records	accuracy 1	accuracy 2	Persons at Home
36	0,23	1.261	90	96	5
59	0,23	1.282	90	95,71	1
33	0,88	4.887	85,6	87,42	1
26	16,39	91.286	88,53	86,62	1
58	0,55	3.036	85,6	84,88	1
32	5,99	33.349	83,85	79,67	1
31	35,48	197.618	86,55	79,42	1
57	0,53	2.942	82,45	76,16	1
56	2,31	12.880	78,28	67,15	2
39	13,33	74.253	88,61	65	2
35	1,69	9.414	81,75	64,34	1
61	1,25	6.940	76,85	63,93	1
50	2,89	16.095	83,14	57	1
38	3,29	18.326	83,23	54,5	2
7	0,91	5.059	78,07	46,35	2
45	4,87	27.104	97,6	41,54	1
34	2,28	12.700	78,63	36,99	1
27	0,16	911	83,02	35,85	1
9	0,22	1.246	79,02	31,71	2
43	2,67	14.857	76,65	28,34	1
54	3,36	18.719	83,12	17,34	2



Selección de la mejor algoritmia a implementar (I)

Estado_Usuario_Codificado = modelo(mes, semana_mes, dia_semana, hora, cuarto_horario, trimestre, estado_meteorológico)

Modelo	min.	Precisión	Parámetros
C5.1	5	77.694	Nivel de Confianza Mínimo por Rama: 25% "Minimal Gain α " = 0,1
Quest	5	64.926	Máxima Profundidad del árbol: 5 "Minimal Gain α " = 0,05
CRT	5	64.212	Máxima Profundidad del árbol: 5 Conjunto de prevención sobreajustado = 30%

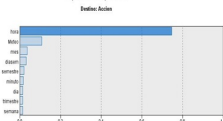
Selección de la mejor algoritmia a implementar (II)

Modelo	min.	Precisión	Parámetros
Red Neuronal	5	61.949	Perceptrón Multicapa, Epsilon = 0,001.
			1 capa intermedia, 3 neuronas
			Ciclos = 500, Learning rate = 0.3
CHAD	5	60.945	Máxima Profundidad del árbol: 5
			Epsilon convergencia CHI = 0,001
			"Minimal Gain α " n = 0,05
Red Bayesiana	5	55.996	Comprobación de independencia: Razón de verosimilitud
			Modelo: Tree Augmented Naïve Bayesian network (TAN)

Selección de la mejor algoritmia a implementar (III)

Mejor Modelo: C5.1 Exactitud (k-fold = 10): 68,12%

Importancia del predictor



accuracy: 68.12% +/- 0.16%

										class
	true S	true P	true E	true T	true O	true C	true D	true U	true X	precision
pred. S	14450	1065	13	0	68	50	1	0	1	92.34%
pred. P	940	3800	695	405	326	1016	52	29	10	52.25%
pred. E	10	320	592	105	41	358	185	0	0	36.75%
pred. T	0	123	177	1566	293	91	201	39	0	62.89%
pred. O	102	296	70	298	1203	242	9	342	28	46.45%
pred. C	44	703	888	40	224	1854	484	85	0	42.90%
pred. D	0	31	175	38	0	138	183	0	0	32.39%
pred. U	0	13	0	73	215	2	0	218	14	40.75%
pred. X	0	0	0	0	1	0	0	2	3	50.00%
class										
recall	92.95%	59.83%	22.68%	62.02%	50.74%	49.43%	16.41%	30.49%	5.36%	

Discretización de eventos en tiempo, duración y frecuencia

Partimos de la base de que el comportamiento y los hábitos humanos se caracterizan por tres atributos de las actividades:

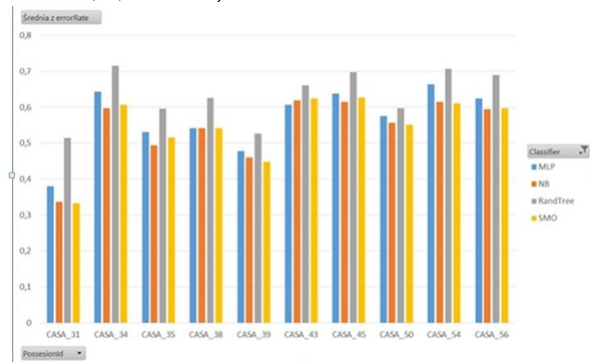
- Tiempo.
- Frecuencia de Duración.
- Frecuencia de Repetición.

Los desvíos en el comportamiento pueden ser identificados mirando los cambios en esos atributos:

- Se calcula la Frecuencia de Duración de los Eventos.
- Eliminamos de los Eventos de memoria del sistema: Frecuencia < 15 minutos.
- Incorporamos al vector de análisis “n” Eventos Anteriores.
- Diferencias con otros trabajos.[Kri14]

Resultado del Modelado

$Evento_Usuario = modelo(mes, dia_semana, día_mes, hora, frecuencia_evento, evento_1, \dots, evento_n)$



El mejor resultado (n=5): 63% de exactitud

Conclusiones Previas

- El indicador “día de la semana” influye mejor que el “día del mes” en la confianza del modelo final.
- El mes produce “overfitting” en los clasificadores.
- Los resultados esperados no dependen en gran medida del volumen de datos.
- Con la ventana deslizante, tomando “n” situaciones anteriores, las predicciones no mejoran.

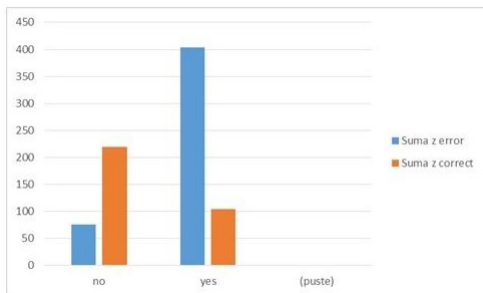
Análisis de series temporales

- Se genera una serie temporal de eventos con una ventana de 24 eventos previos.
- Modelización de la serie temporal mediante la metodología de Box-Jenkins: ARIMA (0.1.3)(1.0.1)
- Coeficiente de determinación $R^2 = 0.18787$

Predicción Binaria sobre Probabilidad de Cambio (I)

Teniendo en cuenta el evento anterior momento que queremos predecir:

- Cuándo el siguiente evento no cambia el predictor tiene éxito.
- En los casos de cambio, el sistema no es capaz de predecir con suficiente confianza cuál va a ser el siguiente evento.



Efectividad de las Reglas Heurísticas

Se revisaron 223 eventos en 29 domicilios:

- 24 eventos, fueron alertas relacionadas con gas, humo o temperatura.
 - Alertas por Gas, (TP 100%).
 - Alertas por Humo (TP 20%)
 - Alertas por Anomalía en Temperaturas (TP 50%).
 - Alertas por Puertas Abiertas (TP 30%).
- 199 eventos relacionados con la actividad.
 - 58 “propuestas de alertas”, es decir, eventos que generaron posibles alerta a revisar por el operador:
 - Actividad Prolongada (TP: 11%)
 - Alertas por Inactividad (TP: 13%)
 - 141 eventos no alertas verdaderas (TN: 70.8%).
- Exactitud: 75%
- Precisión: 12%

Evaluación del Módulo de Detección Automática de Patrones

Se analizan 532 eventos, todos relacionados con la actividad:

- 27 “propuestas de alertas”, es decir, posibles alertas de anomalía sobre los patrones de comportamientos.
 - 16 Verdaderos Positivos (59%).
 - Falsos Positivos:
 - Ocupación de habitaciones no usuales.
 - Salidas y entradas de casa.
- Exactitud: 98%
- Precisión: 59%

Metodología de Control

- C_t : Cambio de Estado del Usuario en el tiempo t .
- C'_t : Predicción del Cambio de Estado del Usuario en el tiempo t .
- $P(C'_t)$: Probabilidad asociada a C'_t .
- E_t : Estado del Usuario en el tiempo t .
- E'_t : Predicción del Estado del Usuario en el tiempo t .
- $P(E'_t)$: Probabilidad asociada a E'_t .
- U_d : Umbral de Confianza para el Domicilio d .

Metodología de Control

Repetir (por cada Domicilio d):

Repetir (para cada instante t):

Calcular $P(C'_t)$ y $P(E'_t)$:

if $P(E'_t) > U_d$ then

if ($true(C'_t)$ and $P(C'_t) > U_d$ and $C'_t \neq C_t$)

then Propuesta_Alerta(t)

elseif ($E'_t \neq E_t$)

then Propuesta_Alerta(t)

endif

endif

else Verdadero_Negativo(t)

endif

Evaluación Global el Sistema

Por cada $Propuesta_Alerta(t)$, el Operador Experto analiza el Contexto de la $Propuesta_Alerta(t)$.

- Si es pertinente \rightarrow Envía Notificación a los sistemas Asistenciales, Anota la $Propuesta_Alerta(t)$ como “VP”.
- Si no es pertinente \rightarrow Anota la $Propuesta_Alerta(t)$ como “FP”.