



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



ESCUELA TÉCNICA
SUPERIOR INGENIEROS
INDUSTRIALES VALENCIA

TRABAJO FIN DE GRADO EN INGENIERÍA BIOMÉDICA

DESARROLLO DE TÉCNICAS DE MEJORA DE CALIDAD DE INFORMACIÓN EN SISTEMAS DE LOCALIZACIÓN A TIEMPO REAL PARA CREACIÓN DE MODELOS DE PREDICCIÓN EN ENTORNOS HOSPITALARIOS

AUTOR: ADRIÁN VICENTE GÓMEZ

TUTOR: JOSÉ LUIS BAYO MONTÓN

COTUTORES: CARLOS FERNÁNDEZ LLATAS, VICENTE TRAVER SALCEDO

Curso Académico: 2018-19

Agradecimientos

Al grupo SABIEN, en especial a Carlos, José Luis y Gema por permitirme realizar este trabajo.

A mis Pitukis, a Anna y a todos aquellos compañeros que hicieron que esta carrera haya merecido la pena del primer día hasta el último por ayudarme durante el camino.

A mi madre, mi hermana, mi padre y mis hermanos por convertirme en quien soy ahora.

Y en especial a David, por ser la luz que me dijo que llegaría muy lejos.

RESUMEN

Los hospitales son entornos de una alta complejidad en los cuales se llevan a cabo diariamente cientos de procesos en los que se ven implicados tanto pacientes como profesionales sanitarios.

El control y seguimiento de estos procesos de manera manual resulta en grandes pérdidas de tiempo por parte de los profesionales que no se ven reflejadas en una mejora directa de la salud de los pacientes. Es por esto por lo que surgen los sistemas de localización a tiempo real (RTLS) los cuales permiten registrar los procesos realizados en el hospital.

Para el aprovechamiento de estos datos serán de gran utilidad las técnicas del *Big Data*, en concreto, la minería de procesos, gracias a la cual se podrán observar los distintos procedimientos que suceden en el hospital. Sin embargo, en ocasiones, la información recopilada por los RTLS contiene datos corruptos que pueden empeorar los resultados de la minería de procesos.

Por ello, en este trabajo se identificaron los errores existentes en los datos de los RTLS y desarrollaron técnicas interactivas para la corrección de estos, pudiendo utilizar los datos corregidos para la generación de modelos de predicción en entornos hospitalarios.

Los modelos generados con los datos corregidos mostraron una mejora en la precisión respecto a los generados con los datos originales.

Palabras clave: Técnicas de mejora de calidad, Corrección de errores, Minería de procesos, Modelos de predicción, Sistemas de localización en tiempo real, Bloque Quirúrgico, Hospital General Universitario de València.

ABSTRACT

Hospitals are highly complex environments where hundreds of processes are performed daily, being implicated both patients and healthcare professionals.

Controlling and tracing these processes manually results in a great waste of time for professionals that is not directly involved on patients' health. Due to that, Real Time Location Systems (RTLS) appear allowing record every process executed in hospitals.

To take advantage of this data, Big Data techniques will be very useful specifically, process mining, thanks to which observe different procedures in hospitals will be possible. However, sometimes, information recorded by RTLS contains corrupt data which could worsen process mining results.

Hence, in this paper there were identified errors presents in RTLS data and interactive techniques for its correction were developed, using the corrected data to generate prediction models in hospital environment.

The prediction models generated with corrected data showed an improvement of precision in contrast with the prediction models generated with the original data.

Keywords: Quality enhancement techniques, Error correction, Process Mining, Prediction models, Surgical service, Hospital General Universitario de València.

RESUM

Els hospitals són entorns amb una alta complexitat en els quals es duen a terme centenars de processos diàriament en els que es veuen implicats tant pacients com professionals sanitaris.

El control i seguiment d'aquests processos de forma manual dóna lloc a grans pèrdues de temps per part dels professionals que no es veuen reflectides en una millora directa de la salut dels pacients. És per això que sorgeixen els sistemes de localització a temps real (RTLS) els quals permeten registrar els processos realitzats a l'hospital.

Per a l'aprofitament d'aquestes dades seran de gran utilitat les tècniques del *Big Data*, concretament, la mineria de processos, gràcies a la qual es podran observar els diferents procediments que es duen a terme a l'hospital. No obstant això, en ocasions, la informació recopilada per els RTLS conté dades corruptes que podrien empitjorar els resultats de la mineria de processos.

Per aquest motiu en aquest treball es van identificar els errors existents en les dades dels RTLS i es van desenvolupar tècniques interactives per a la correcció d'aquests, podent utilitzar les dades corregides per a la generació de models de predicció en entorns hospitalaris.

Els models generats amb les dades corregits van mostrar una millora en la precisió respecte als generats amb les dades originals.

Paraules claus: Tècniques de millora de la qualitat, Correcció de errors, Minería de Processos, Models de predicció, Sistemes de localització a temps real, Bloc quirúrgic, Hospital General Universitario de València.

ÍNDICE GENERAL

DOCUMENTOS CONTENIDOS EN EL TFG

- Memoria
- Presupuesto

MEMORIA

Desarrollo de Técnicas de Mejora de Calidad de Información en
Sistemas de Localización a Tiempo Real para Creación de Modelos de
Predicción en Entornos Hospitalarios

Adrián Vicente Gómez

Índice de la Memoria

CAPÍTULO 1. INTRODUCCIÓN	1
1.1 Objetivos.....	3
CAPÍTULO 2. MARCO TEÓRICO.....	5
2.1 Sistemas de Localización en Tiempo Real	5
2.2 Minería de Procesos en RTLS	8
2.3 Modelos de Predicción: Modelos Ocultos de Markov	11
2.3.1 Precedentes.....	11
2.3.2 Modelos ocultos de Márkov	11
CAPÍTULO 3. MATERIAL Y MÉTODOS.....	13
3.1 La Base de Datos	13
3.2 El Hospital General Universitario de Valencia	14
3.2.1 Organización del bloque quirúrgico del HGUV	15
3.2.2 Flujos de Trabajo en la Unidad Quirúrgica del HGUV	20
3.3 Aplicación de Minería de Procesos y Generación de Modelos	21
3.3.1 Creación de HMM.....	22
3.3.2 Corrección de errores	22
3.4 Identificación de Errores en los Datos.....	23
3.5 Técnicas de Mejora de la Calidad de la Información	24
3.6 Uso de la Corrección de Errores del PMapp	28
3.7 Creación de Modelos de Predicción y Validación.....	29
3.7.1 Entrenamiento de modelos	29
3.7.2 Validación de los modelos	30
CAPÍTULO 4. RESULTADOS.....	31
4.1 Mejora de la Calidad de los Datos.....	31
4.2 Resultados de la Predicción	32
4.2.1 Predicción del estado Preparation	33
4.2.2 Predicción del estado Intervention	34
4.2.3 Predicción del estado Recovery	36

4.2.4	Predicción del estado In dressing room.....	37
4.2.5	Predicción del estado Adaptation.....	39
CAPÍTULO 5. DISCUSIÓN		41
5.1	Evaluación del Efecto de las TMC en la Calidad de los Datos	41
5.2	Evaluación del Efecto de las TMC en la Predicción	41
CAPÍTULO 6. CONCLUSIONES		43
6.1	Líneas Futuras.....	43
CAPÍTULO 7. REFERENCIAS.....		45

Índice de Ilustraciones

Ilustración 1.1 Evolución de personas en lista de espera quirúrgica en España.	2
Ilustración 2.1. Arquitectura del sistema RTLS de MYSPHERA®.	5
Ilustración 2.2. Interfaz de la aplicación MYHOSPITAL..	6
Ilustración 2.3 Ejemplo del panel informativo para familiares. a.	7
Ilustración 2.4. Ejemplo del panel informativo para el personal.	8
Ilustración 2.5. Minería de procesos en salud con registros de RTLS.	9
Ilustración 2.6. Ejemplo de efecto espagueti.....	10
Ilustración 2.7. Representación de un HMM.	11
Ilustración 3.1. Fotografía de la entrada de HGUV..	15
Ilustración 3.2. Plano de la planta baja del bloque quirúrgico del HGUV.....	16
Ilustración 3.3. Plano de la segunda planta del bloque quirúrgico de HGUV.	16
Ilustración 3.4. Zona de recepción del área quirúrgica de HGUV.	17
Ilustración 3.5. Área prequirúrgica del HGUV.	18
Ilustración 3.6. Quirófano del HGUV.	18
Ilustración 3.7. Área postquirúrgica del HGUV.	19
Ilustración 3.8. Vestuario del HGUV.....	19
Ilustración 3.9. Flujo de trabajo ideal de un paciente en CSI.....	20
Ilustración 3.10. Flujo de trabajo ideal de un paciente para CE.....	20
Ilustración 3.11. Flujo de trabajo ideal de un paciente para HOSP..	21
Ilustración 3.12 Pantalla principal del PMAApp.	21
Ilustración 3.13. Ventana de diseño de TPAs del PMAApp.....	23
Ilustración 3.14. Ejemplo de traza con reflexión.	24
Ilustración 3.15. Aspecto de la salida de load_data.m.	25
Ilustración 3.16. Aspecto de una traza correspondiente a un campo de la salida de load_data.m.	25
Ilustración 3.17. Diagrama de flujo de la función fix_ref.m.....	26
Ilustración 3.18. Diagrama de flujo de la función fix_desas.....	27
Ilustración 3.19. TPA para corpus original, corpus tras TMC y corpus de CSI.	28
Ilustración 3.20. TPA para corpus de CCI.	28

Ilustración 3.21. Funcionamiento del proceso de validación.	30
Ilustración 4.1. Resultado de la aplicación del algoritmo de MP al corpus original.....	31
Ilustración 4.2. Resultado de la aplicación del algoritmo de MP al corpus tras las TMC.	32
Ilustración 4.3 Diagrama de caja y bigotes con los resultados de la predicción para el estado Preparation.	34
Ilustración 4.4. Diagrama de caja y bigotes con los resultados de la predicción para el estado Intervention.	35
Ilustración 4.5. Diagrama de caja y bigotes con los resultados de la predicción para el estado Recovery.....	37
Ilustración 4.6. Diagrama de caja y bigotes con los resultados de la predicción para el estado In dressing room.	38
Ilustración 4.7. Diagrama de caja y bigotes con los resultados de la predicción para el estado Adaptation.	40

Índice de Tablas

Tabla 1.1 Gasto sanitario total según proveedor de atención de salud en España, año 2015	1
Tabla 1.2. Número de operaciones y coste estimado de las misma en España, año 2016.	2
Tabla 3.1. Valores del campo State.....	13
Tabla 3.2. Valores del campo Program (izquierda) y su significado (derecha).	14
Tabla 3.3. Número de trazas totales, para entrenamiento y para validación en los distintos corpus utilizados.....	29
Tabla 4.1. Predicción del estado Preparation con los distintos corpus en minutos.....	33
Tabla 4.2. Predicción del estado Intervention con los distintos corpus en minutos.....	35
Tabla 4.3. Predicción del estado Recovery con los distintos corpus en minutos.	36
Tabla 4.4. Predicción del estado In dressing room con los distintos corpus en minutos.	38
Tabla 4.5 Predicción del estado Adaptation con los distintos corpus en minutos.	39

CAPÍTULO 1. INTRODUCCIÓN

La atención a la salud de la población es uno de los campos que más actividad genera, tanto es así que en España conlleva un gasto total de 101 millones de euros (contando sector privado y público), lo que representa un 9,1 % del Producto Interior Bruto (PIB)[1]. Este presupuesto es destinado a diversas actividades como gasto farmacéutico, atención a domicilio o cuidados de larga duración. Una de las actividades que más recursos consume son los hospitales, como se puede observar en la tabla 1.1.

*Tabla 1.1 Gasto sanitario total según proveedor de atención de salud en España, año 2015.
Fuente: [2]*

Proveedor	Gasto (millones de euros)
Hospitales	41.897
Establecimientos de atención medicalizada y residencial	5.562
Proveedores de atención ambulatoria	25.786
Minoristas y otros proveedores de productos médicos	21.944
Suministro y administración de programas de salud pública	695
Administración general de la salud y los seguros médicos	3.002
Otras ramas de actividad	1.079
Organizaciones Sanitarias Internacionales	9
Total	99.974

Los hospitales son estructuras de alta complejidad, que se dividen en múltiples bloques y servicios para asegurar un mejor funcionamiento. Además de esto, en los últimos tiempos sus presupuestos se han visto ajustados por motivos de la crisis económica, a la vez que el envejecimiento de la población provoca que más personas necesiten habitualmente de sus servicios.[3]

De entre todos los servicios, uno de los más relevantes y, por tanto, susceptible de ser optimizado es el bloque quirúrgico. La importancia de este servicio radica en dos puntos:

- Se trata del área que más costes genera en un hospital[4]. En el año 2016 en España se realizaron más de 3 millones y medio de operaciones que con los datos de la tabla 1.2, se puede deducir que supone un 20% del gasto destinado a hospitales de la tabla 1.1.

Tabla 1.2. Número de operaciones y coste estimado de las misma en España, año 2016. Fuente: elaboración propia en base a [5].

Área de Actividad	Actos	Coste medio estimado (euros)	Coste estimado (millones de euros)
Total Actos Quirúrgicos	2.610.289	-	8.766
Con hospitalización	1.411.091	6.143	8.668
Con Cirugía Mayor Ambulatoria	1.199.198	823	98

- Se trata del área que genera listas de esperas más largas. En España, la lista de espera quirúrgica no ha parado de crecer en los últimos años como se puede observar en la ilustración 1.1, alcanzando en el primer semestre de 2018 las 584 mil personas a la espera de una intervención, con 93 días de espera media[6].

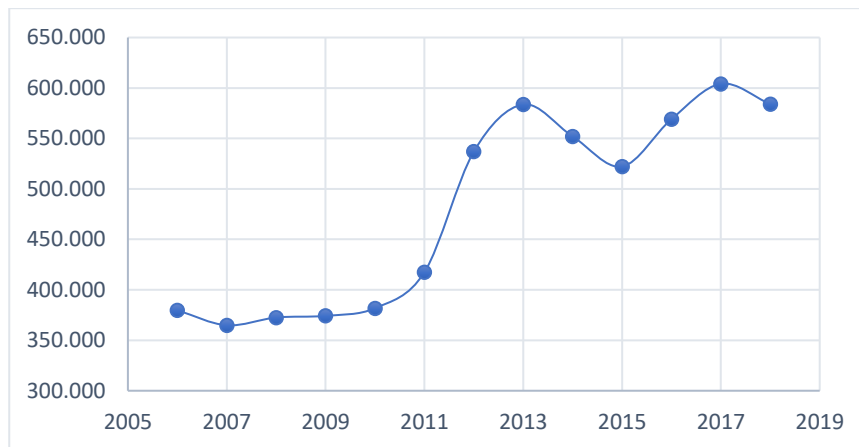


Ilustración 1.1 Evolución de personas en lista de espera quirúrgica en España. Fuente: elaboración propia con datos de [6].

Por tanto, mejorar el funcionamiento de los procesos quirúrgicos podría reducir el gasto total en sanidad, además de permitir la realización de un mayor número de intervenciones, reduciéndose así las listas de espera.

La optimización de los procesos intrahospitalarios en general, y de los procesos quirúrgicos en particular, resulta una necesidad imperante para poder ofrecer una mejor atención con el mínimo coste posible. Pero si se quiere realizar es necesario conocer el funcionamiento de los procesos que suceden en los centros de atención sanitaria, y esto es podrías ser posible a través los registros del sistema informático del hospital (HIS). Estos registros, sin embargo, no siempre presentan la calidad adecuada, generando ruido que no permite el correcto aprovechamiento de los datos. Para solucionar este problema, muchos hospitales están comenzando a utilizar sistemas de localización en tiempo real (RTLS), que permiten mantener un registro de manera más automatizada reduciendo la posibilidad de fallo humano [3], [7]. A pesar de que estos sistemas representan un gran avance de cara a la optimización de procesos hospitalarios, debido a limitaciones del propio dispositivo y a que el factor humano sigue estando presente, los registros generados por RTLS pueden presentar problemas de calidad que reducen su fiabilidad de cara a utilizarlos como fuente de datos.

1.1 OBJETIVOS

El objetivo principal de este trabajo es desarrollar técnicas de mejora de calidad (TMC) de la información generada por RTLS para así poder generar modelos de predicción intrahospitalarios más precisos. Este objetivo puede desglosarse en tres puntos:

- Identificar los posibles errores que reduzcan la calidad de los datos.
- Desarrollar técnicas interactivas para detectar y corregir los errores identificados, manteniendo el mayor número de trazas posibles.
- Evaluar el efecto de las TMC en la generación de modelos de predicción.

Además de este objetivo, durante el proceso se abordarán los siguientes objetivos específicos:

- Profundizar en el conocimiento del funcionamiento de RTLS.
- Utilizar la minería de procesos para analizar grandes cantidades de datos.
- Comprender el funcionamiento de bloque quirúrgico del hospital.
- Generar modelos de predicción y validarlos mediante herramientas no propias.

CAPÍTULO 2. MARCO TEÓRICO

2.1 SISTEMAS DE LOCALIZACIÓN EN TIEMPO REAL

Los RTLS son una tecnología que permite realizar una trazabilidad de objetos o personas basándose en un sistema compuesto por:

- Dispositivos emisores de una señal que se encuentran integrados en objetos como tarjetas, pulseras o collares.
- Dispositivos receptores de la señal que permiten identificar la posición del objeto emisor.[3]

Este tipo de sistemas ha sido aplicado con éxito en múltiples hospitales, permitiendo realizar un seguimiento tanto de equipos como de pacientes sin necesidad de dedicar tiempo del personal a realizar un registro manual del tiempo que pasan en cada localización.

Una de las soluciones de RTLS aplicado en salud que más impacto está teniendo en la actualidad es la implementada por la empresa valenciana MYSPHERA®, que en la actualidad se encuentra instalada en más de 50 hospitales de Europa. Esta solución está implementada con la arquitectura que se observa en la ilustración 2.1. A continuación se realizará una descripción del funcionamiento de cada una de sus partes [7].

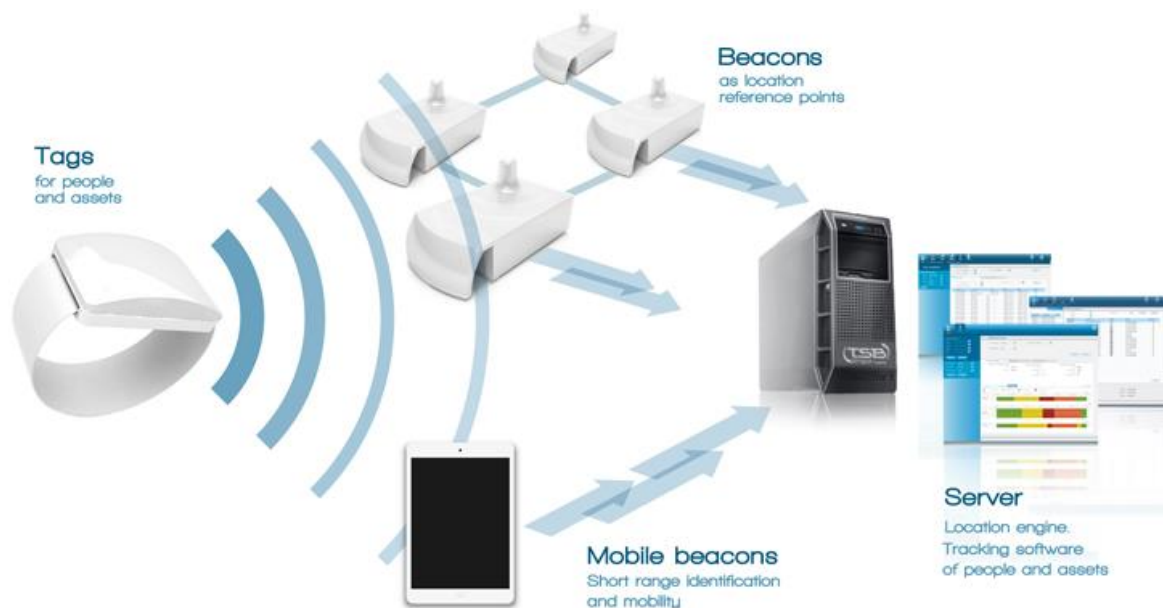


Ilustración 2.1. Arquitectura del sistema RTLS de MYSPHERA®. Fuente: [8]

- Los **tags** son asignados a una persona u objeto. Desde el momento en que son asignados comienzan a emitir una señal periódicamente mediante tecnología ZigBee, esta señal será captada por las balizas lo que permitirá la localización de los tags.

El uso de tecnología ZigBee permite que el consumo de las pulseras sea reducido, consiguiendo alargar la vida de las baterías entre dos y tres años.

- Las **balizas** se encuentran situadas en localizaciones específicas del edificio en el cual se realiza la instalación del sistema RTLS. Estas balizas reciben la señal enviada por las pulseras y registran en el servidor de la instalación la intensidad con la que llega esa señal.
- Las **balizas móviles** son dispositivos portátiles (teléfonos inteligentes o tabletas) que permiten la identificación de una pulsera en concreto. Se utilizan en conjunto con los lectores de escritorio, los cuales permiten la asignación del tag a un dispositivo o a un paciente.
- El **servidor** de MYS-PHERA[®] se encuentra normalmente instalado de manera física en el hospital en que se realiza la instalación. Este servidor cuenta con dos módulos principalmente:
 - **SPHERAONE**. Es el programa que ejecuta el algoritmo que permite identificar la localización del tag gracias a los datos enviados por las balizas. Con este programa también se puede comprobar el estado de los distintos dispositivos del sistema RTLS.
 - **MYHOSPITAL**. Es el programa que permite realizar las asignaciones y el seguimiento de los tags, así como la visualización de estadísticas de los procesos realizados en el hospital que permiten a los gestores conocer el estado del centro. La interfaz de esta aplicación puede observarse en la ilustración 2.2.

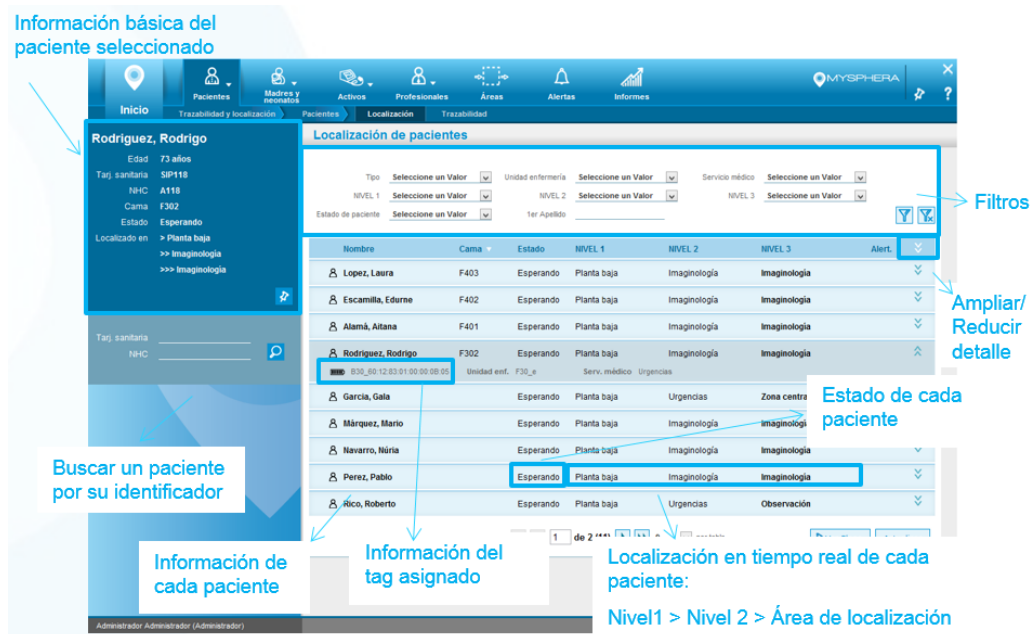


Ilustración 2.2. Interfaz de la aplicación MYHOSPITAL. Todos los datos de pacientes que aparecen nos pertenecen pacientes reales si no a pacientes de ejemplo. Fuente: cedido por MYS-PHERA[®].

Entre los servicios que la solución de MYSPHERA® ofrece a los hospitales se puede distinguir entre tres grupos:

- **Información para familiares de pacientes.** Se proporciona información al familiar del paciente sobre el estado asociado a la ubicación en la que se encuentra el paciente a través de un código anónimo proporcionado durante la recepción al paciente que aparece en una pantalla en la sala de espera, similar a la de la ilustración 2.3. De este modo, por ejemplo, en un servicio de urgencias, el familiar puede saber si el paciente se encuentra en triaje, en tratamiento... Evitándose así interrupciones al personal, reduciendo la ansiedad de los familiares debido al desconocimiento del estado del paciente y aumentando la satisfacción en general [8].



Ilustración 2.3 Ejemplo del panel informativo para familiares. Arriba cirugía en pacientes hospitalizados y abajo cirugía mayor ambulatoria. Fuente: cedido por MYSPHERA®.

- Información para el personal.** Se proporciona una información más completa que la que recibe el familiar, como se puede observar en la ilustración 2.4 en este panel se puede observar el nombre real del paciente, su localización exacta, el tipo de proceso al que está siendo sometido, la especialidad a la que corresponde, el estado, el tiempo que se ha mantenido en cada estado y el tiempo total. De esta forma es posible optimizar los recursos del hospital y agilizar el servicio, evitando, por ejemplo, los cuellos de botella.

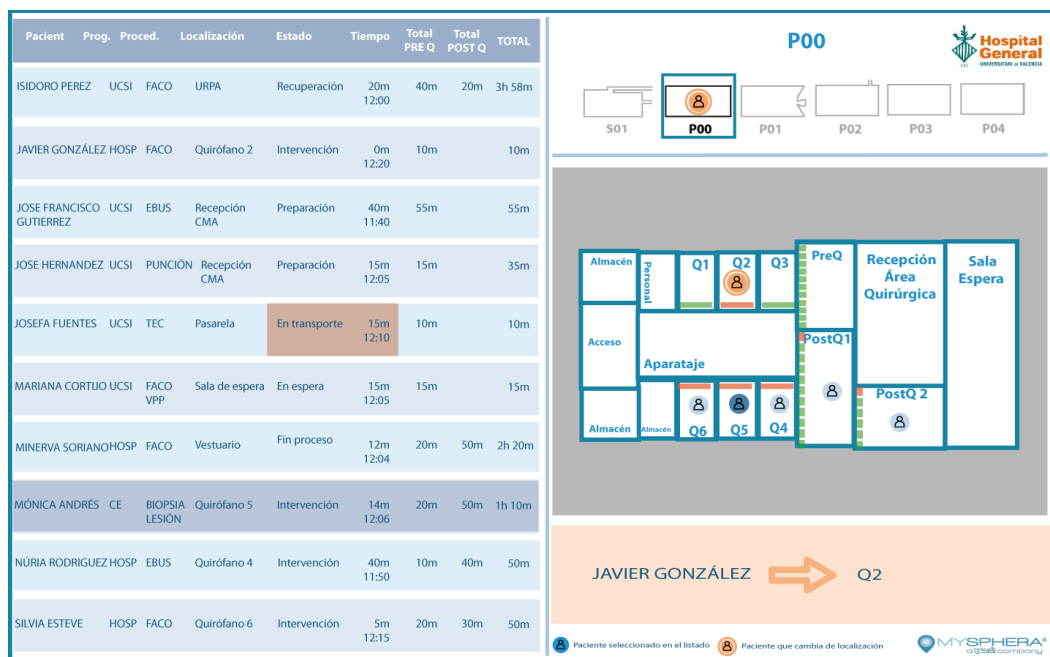


Ilustración 2.4. Ejemplo del panel informativo para el personal. Todos los pacientes que aparecen no son pacientes reales si no pacientes de ejemplo. Fuente: cedido por MYSOPHERA®.

- Información para los gestores.** Como ya se ha indicado antes, los gestores pueden visualizar informes periódicos con estadísticas del servicio como tiempos en la sala de espera, ocupación de consultas o número de pacientes tratados. Además, también permite mantener un control del inventario de manera automático.

Como se puede observar, la cantidad de datos generada por los sistemas RTLS sugiere que para estudiar correctamente los procesos que suceden en un hospital sea necesario el uso de técnicas para el análisis de datos masivos como la minería de procesos.

2.2 MINERÍA DE PROCESOS EN RTLS

La minería de procesos (MP) es un conjunto de técnicas a través de las cuales se pueden construir modelos comprensibles por los seres humanos a partir de registros de eventos, que en MP se suele llamar *log* o *corpus*. A partir de estos modelos es posible analizar visualmente los procesos que existen de manera subyacente en los registros[9].

Estos procesos son representados mediante flujos de trabajos en los que los nodos representan una ubicación, un estado o una actividad; y las flechas que los unen representan el orden en el que se pasa de un nodo a otro. Resulta evidente que, si el registro del que disponemos contiene datos como los que se obtienen mediante un sistema RTLS, la MP se convierte en una herramienta de gran utilidad para generar valor añadido a los datos.

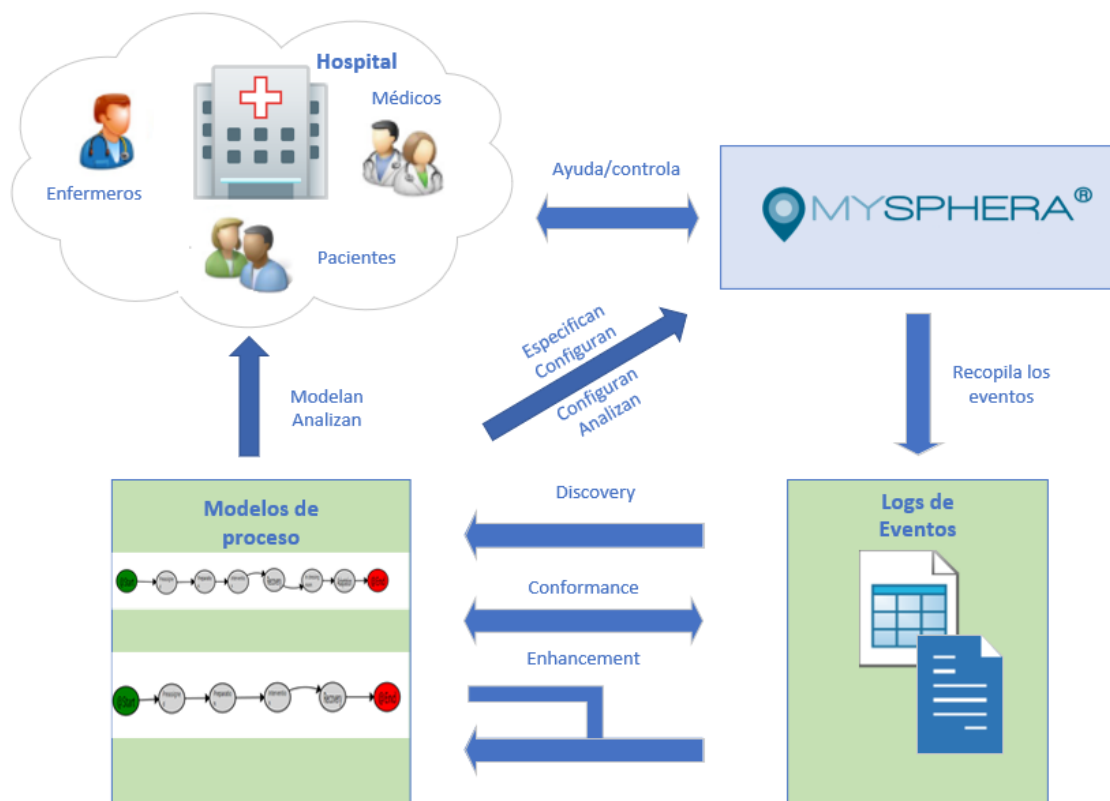


Ilustración 2.5. Minería de procesos en salud con registros de RTLS. Fuente: elaboración propia basada en [9] y [10].

Como se observa en la figura 2.5, existen tres tipos de MP [9]:

- *Discovery*. Es una técnica que permite descubrir los flujos de trabajo existentes en los datos sin ninguna información a priori. Una aplicación de este tipo de técnica con datos de RTLS es la caracterización del comportamiento de individuos en residencias de la tercera edad[11].
- *Conformance*. Es una técnica que permite observar cómo los datos presentes en el registro se ajustan a un modelo teórico predefinido. Con este tipo de MP se puede evaluar si los pacientes están siguiendo las rutas definidas originariamente por el centro.

- *Enhancement.* Es una técnica que utiliza el modelo descubierto con los datos contenidos en el corpus para reparar o mejorar el modelo teórico de flujo de trabajo que debería seguirse, generando así un nuevo modelo. Con este tipo de MP se puede observar si existen cuellos de botella y modificar la rutina del servicio para evitarlos.

Una de las limitaciones de la MP es que, ante bases de datos con mucho ruido, aparece el denominado efecto espagueti similar al de la ilustración 2.6, el cual provoca que el modelo creado sea incomprensible, perdiéndose así una de las principales ventajas de la MP[7]. Por tanto, la reducción del ruido en la base de datos debe ser uno de los primeros pasos a seguir antes de utilizar la MP.

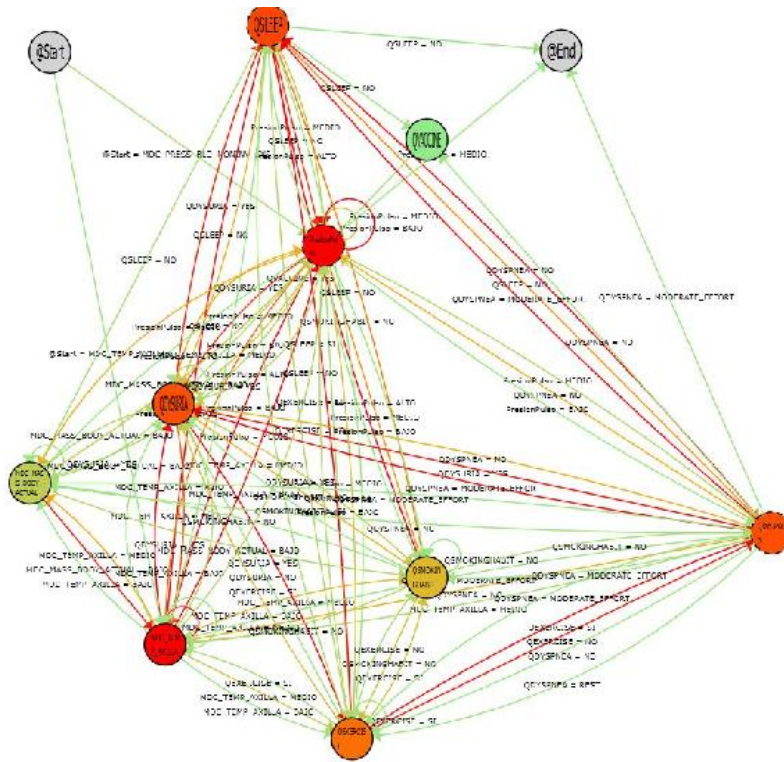


Ilustración 2.6. Ejemplo de efecto espagueti. Fuente: [12].

Durante el presente trabajo se utilizarán los dos primeros tipos de MP para detectar los posibles errores que aparezcan en corpus que contengan información de RTLS, además de para generar flujos que puedan ser utilizados por los modelos de predicción.

2.3 MODELOS DE PREDICCIÓN: MODELOS OCULTOS DE MARKOV

2.3.1 Precedentes

Tradicionalmente, la planificación de intervenciones en los hospitales ha dependido del cirujano, el cuál determinaba cuánto tiempo estimaba que era necesario para realizar la operación en base a su experiencia. Este tipo de planificación puede inducir en una gran variedad de errores, como la sobreestimación de la duración de la operación, lo cual conduciría a un despilfarro de tiempo y recursos por parte del hospital.

Debido a esto, desde la incorporación de los registros electrónicos a la rutina diaria de los centros médicos, múltiples autores han propuesto la utilización de estos datos para la creación de modelos predictivos que permitan optimizar los procesos en el hospital [13], [14].

Estos modelos de predicción en entornos hospitalarios se han elaborado a través de diferentes técnicas: análisis de la varianza (ANOVA)[13]; redes neuronales recurrentes (RNN)[14] o modelos ocultos de Márkov (HMM) [15].

Debido a las características de los datos obtenidos por RTLS, que coinciden con una sucesión de eventos, se ha decidido que los HMM son la mejor opción para generar modelos de predicción.

2.3.2 Modelos ocultos de Márkov

Los HMM son modelos estocásticos que permiten modelar secuencias de eventos en las cuales el último estado solamente depende del estado actual del sistema, de tal forma que los estados anteriores al presente estado no tienen impacto en el futuro.

Se denominan modelos ocultos puesto que asumen que existen dos tipos de eventos: los observables y los ocultos. Si se observa la ilustración 2.7, los círculos azules se corresponden con los estados observables, y los círculos blancos son estados no observables que influyen en las probabilidades de transición de un estado a otro y que han de ser modelados[16].

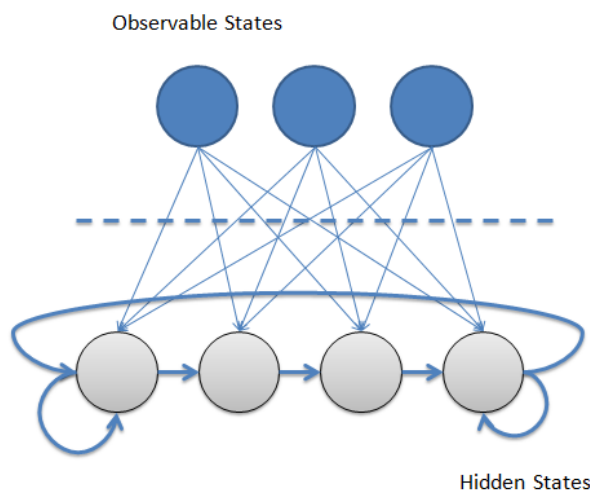


Ilustración 2.7. Representación de un HMM. Fuente: [17].

Tradicionalmente, los HMM han sido caracterizados por tres problemas fundamentales [18]:

1. Dados los parámetros de un HMM y una secuencia de observaciones, determinar la probabilidad de que ocurra la secuencia de observaciones. Este problema se resuelve mediante el algoritmo *Forward-Backward*.
2. Dada una observación y los parámetros de un HMM, encontrar la secuencia oculta que mejor explique la observación. Este problema se resuelve mediante el algoritmo de Viterbi.
3. Dada una secuencia y un conjunto de estados en un HMM, obtener los parámetros del modelo. Este problema se resuelve mediante el algoritmo de Baum-Welch.

Dado que gracias a la MP es posible conocer la secuencia intrínseca en los datos de los que se dispone, bastará con utilizar el algoritmo *Forward-Backward* para determinar las probabilidades asociadas. Sin embargo, se ha de tener en cuenta que el número de datos utilizados para generar los modelos debe ser elevado y presentar variabilidad en el proceso para así evitar el riesgo de obtener modelos solo válidos para el conjunto de entrenamiento y no compatibles con datos nuevos, es decir, evitar el sobreentrenamiento.

CAPÍTULO 3. MATERIAL Y MÉTODOS

3.1 LA BASE DE DATOS

El conjunto de datos con el que se ha trabajado ha sido cedido por la empresa MYSPHERA®. Se trata de registros correspondientes al servicio de cirugía del Hospital General Universitario de Valencia (HGUV) desde el 1 de diciembre de 2017 hasta el 28 de diciembre de 2018, donde todos los datos que permitieran la identificación del paciente han sido anonimizados.

Se ha dispuesto de la siguiente información asociada a cada registro:

- **Tag.** Es el identificador de la pulsera localizadora. Se trata de un código de 12 caracteres que puede incluir números y letras.
- **Trace.** Es el identificador del paciente que entra en el servicio quirúrgico. Se trata de un identificador de 5 números que no permite la identificación de la identidad real del paciente.
- **State.** Es el estado dentro del proceso quirúrgico en el que se encuentra el paciente, se distingue entre once estados diferentes que se pueden ver en la tabla 3.1, estos estados se corresponden en su mayoría con las zonas que se explicarán en siguiente apartado.

Tabla 3.1. Valores del campo State. Fuente: elaboración propia.

Estado	
Inicio BQ	Adaptation
Preassigned	In transit
Preparation	ICU
Intervention	UNKNOWN
Recovery	NULL
Dressing Room	

- **Program.** Es el programa de cirugía dentro del cual se engloba la operación a la que va a ser sometido el paciente, en el corpus aparecen cinco programas distintos que se pueden ver en la tabla 3.2.

Tabla 3.2. Valores del campo Program (izquierda) y su significado (derecha). Fuente: elaboración propia.

Programa	Significado
CSI	Cirugía Sin Ingreso
CE	Cuida Enlace
HOSP	Hospitalización
UNKNOWN	Desconocido
NULL	No se dispone de información sobre el programa

- **Medical_Service.** Es el servicio médico al que se asocia la intervención a realizar, se distingue entre 25 servicios diferentes. Se representa mediante 3 letras. Ejemplo: oftalmología (OFT), urología (URO), ginecología (GIN).
- **Surgeon.** Es el identificador del cirujano encargado de la operación. Se trata de un código numérico que representa al cirujano sin desvelar ningún dato personal del mismo.
- **Surgical_Process.** Es el proceso quirúrgico al que va a ser sometido el paciente, en el período del que se dispone se realizaron 982 procesos quirúrgicos distintos. Se trata de una variable en la que aparece el nombre completo de la intervención.
- **In.** Es la fecha y hora en la que el paciente comienza un estado. Es de la forma dd/mm/aaaa hh:mm:ss.
- **Out.** Es la fecha y hora en la que el paciente termina un estado. Es de la forma dd/mm/aaaa hh:mm:ss.
- **Duration.** Es el tiempo transcurrido en un estado desde In hasta Out. Es de la forma hh:mm:ss.

En total en el corpus original de datos se dispone de 18.481 trazas diferentes que representan, teóricamente, el número de pacientes que han pasado por el servicio quirúrgico del hospital.

3.2 EL HOSPITAL GENERAL UNIVERSITARIO DE VALENCIA

El HGUV, visible en la ilustración 3.1, es un centro médico fundado hace más de 500 años que cumple una función tanto asistencial, como docente e investigadora [19]. Se trata del hospital de referencia del área de salud del departamento de València-Hospital General, al cual correspondían, en el año 2018, 360.488 pacientes a los que hay que añadir aquellos que se atienden a pesar de pertenecer a otro departamento. Esta cantidad de pacientes genera una gran actividad asistencial por parte del HGUV, que en 2018 atendió más de 140 mil urgencias y realizó en torno a 28 mil intervenciones [20].

Por ello, en su constante búsqueda de mejora, en el año 2015, el HGUV implantó la solución de RTLS de la empresa MYSPHERA® en sus servicios de urgencias y cirugía. En este trabajo, como se indicó en el apartado 3.1, se han utilizado únicamente datos del bloque quirúrgico, la organización del cuál se explicará en el siguiente apartado.



Ilustración 3.1. Fotografía de la entrada de HGUV. Fuente:[21].

3.2.1 Organización del bloque quirúrgico del HGUV

El bloque quirúrgico de un hospital es el área física destinada a la realización de intervenciones quirúrgicas dentro de este. Además de los quirófanos también forma parte del bloque la zona de admisiones, la prequirúrgica y la postquirúrgica. En el caso del HGUV inauguró un nuevo bloque quirúrgico en el año 2015, incluyendo la instalación la solución de RTLS de MYSPHERA®.

Tras la realización del despliegue fue necesario definir unas áreas que representaran la localización de las balizas dentro del bloque, y para esto se definen dos niveles de agregación: localización y estado. A pesar de que el bloque quirúrgico cuenta con 5 plantas, en las que se incluye el sótano, para este trabajo solo resultan de interés la planta baja y la segunda planta, por tanto, no se detallarán los despliegues realizados en el resto de las plantas. Los planos de la planta baja y la segunda se corresponden con las ilustraciones 3.2 y 3.3, en la primera de ellas se realizan las operaciones de cirugía mayor ambulatoria (CMA) y en la segunda aquellas que requieren hospitalización.

Desarrollo de Técnicas de Mejora de Calidad de Información en Sistemas de Localización a Tiempo Real para Creación de Modelos de Predicción en Entornos Hospitalarios

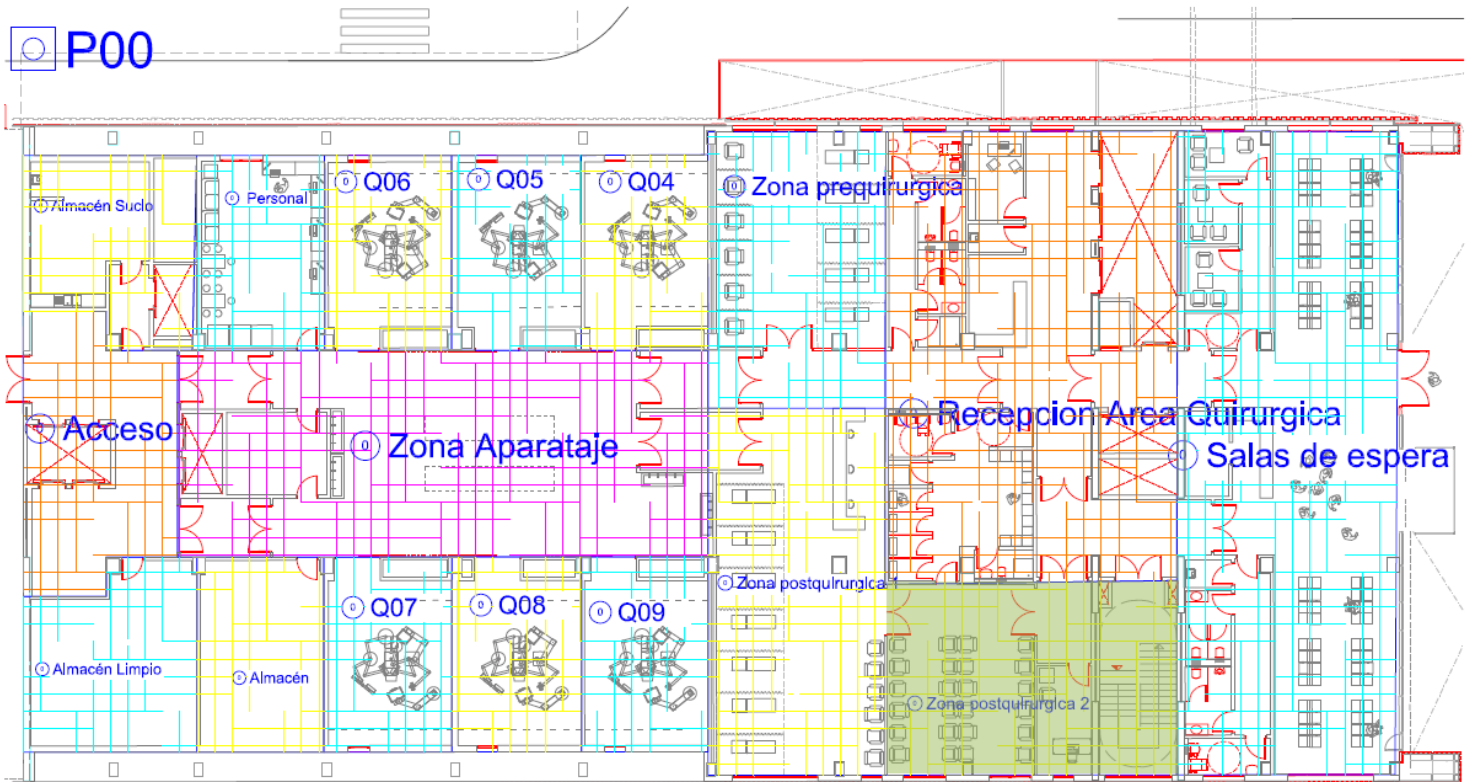


Ilustración 3.2. Plano de la planta baja del bloque quirúrgico del HGUV. Fuente: cedida por MYSHERA®.

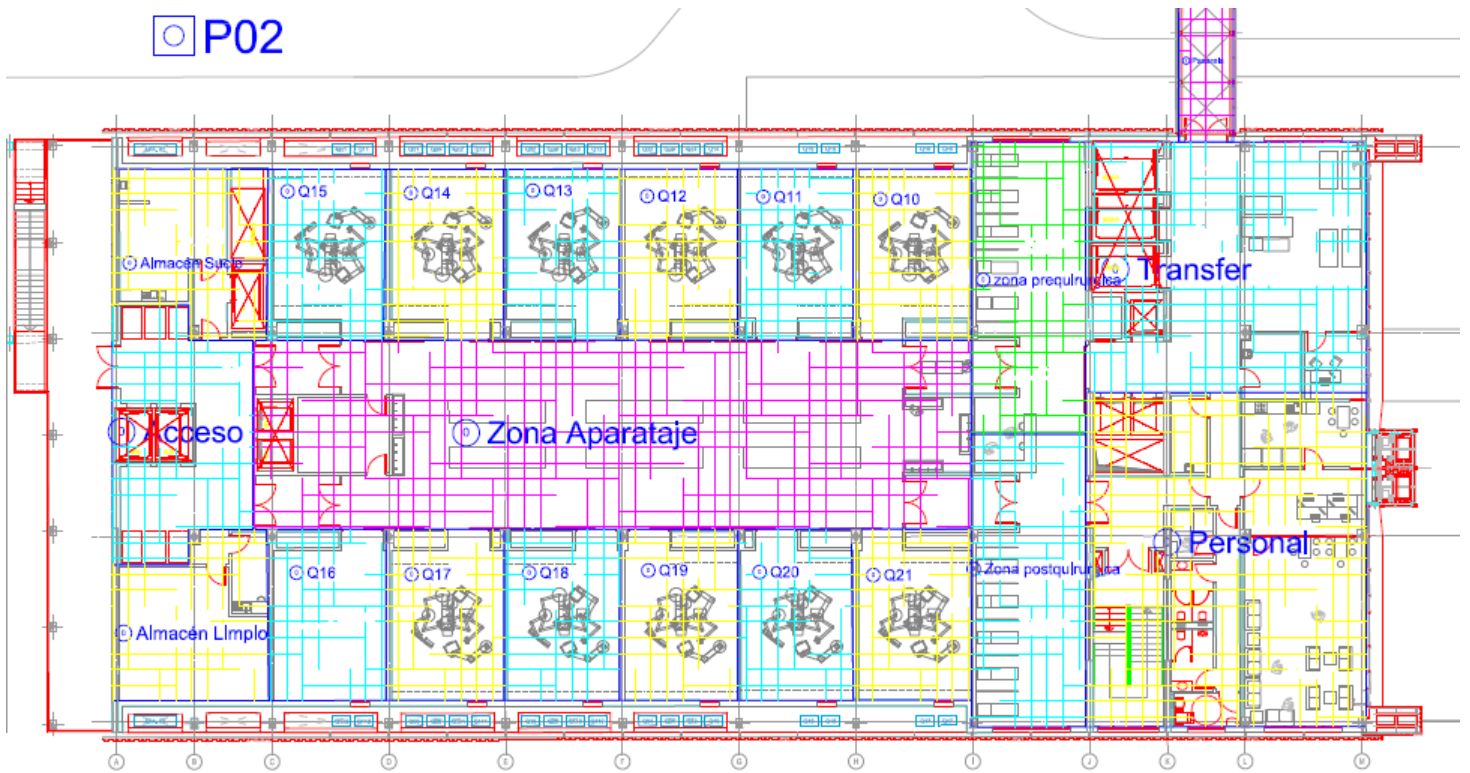


Ilustración 3.3. Plano de la segunda planta del bloque quirúrgico de HGUV. Fuente: cedida por MYSHERA®.

- **Área asociada a una localización.** Se corresponde a la localización en la que está localizada la baliza, como podría ser el quirófano concreto en el que se realiza la intervención, la zona de aparataje o la sala de espera.
- **Área asociada a un estado.** Se dividió el bloque quirúrgico en tantas áreas como estados aparecen en la tabla 3.1 exceptuando *NULL*, *Unknown*, ya que se tratan de estados que no representan localizaciones reales. Estas áreas, son:
 1. *Preasignación.* Se corresponde con el estado *Preassigned* de la base de datos. En el plano de la ilustración 3.2 es la zona en la que se encuentran la recepción del área quirúrgica y las salas de espera. En esta zona (ilustración 3.4), el personal encargado asigna cada pulsera a cada paciente el día anterior a la intervención si esta se encuentra programada, en caso de que la intervención se encuentre programada para un lunes, se asignará la pulsera el último día laboral anterior a la intervención.



Ilustración 3.4. Zona de recepción del área quirúrgica de HGUV. Fuente: cedida por MYSHERA®.

2. *Preparación.* Se corresponde con el estado *Preparation* de la base de datos. Tanto en el plano de la segunda como de la planta baja es la zona marcada como zona prequirúrgica. Esta zona (ilustración 3.5) es a la que acceden los pacientes antes de ser intervenidos y donde son monitorizados, preparándolos para la intervención.



Ilustración 3.5. Área quirúrgica del HGUV. Fuente: cedida por MYSHERA®.

3. *Intervención.* Se corresponde con el estado *Intervention* de la base de datos. En ambos planos comprende tanto los quirófanos como la zona de aparataje por donde se accede a los mismos (ilustración 3.6). En esta zona se realizará la intervención.



Ilustración 3.6. Quirófano del HGUV. Fuente: cedida por MYSHERA®.

4. *Recuperación.* Se corresponde con el estado *Recovery* de la base de datos. En el plano de la planta baja se corresponde con la zona posquirúrgica 1 y en el de la segunda planta con la zona postquirúrgica. En esta zona (ilustración 3.7) todos los pacientes pasan un tiempo después de la operación mientras se recuperan de la anestesia.



Ilustración 3.7. Área postquirúrgica del HGUV. Fuente: cedida por MYSHERA®.

5. *Vestuario*. Se corresponde con el estado *In dressing room* de la base de datos. No se observa en los planos. En esta zona (ilustración 3.8) los pacientes de CMA, tras la recuperación se cambian de ropa antes de pasar al siguiente estado.



Ilustración 3.8. Vestuario del HGUV. Fuente: cedida por MYSHERA®.

6. *Adaptación*. Se corresponde con el estado *Adaptation* de la base de datos. En el plano de la planta baja es la zona postquirúrgica 2. En esta zona los pacientes de CMA descansan un período de tiempo antes de poder ser dados de alta y después de haberse cambiado de ropa.
7. *Transfer*. Se corresponde con el estado *In transit* de la base de datos. Es un estado exclusivo de los pacientes que se encuentran ingresados puesto que es el estado en el que se mantiene la pulsera mientras un celador transporta al paciente desde su habitación hasta el bloque quirúrgico. En el plano de la segunda planta está identificado como *Transfer*.

8. **UCI.** Se corresponde con el estado ICU de la base de datos. Es la unidad de cuidados intensivos y se encuentra en la cuarta planta del bloque quirúrgico, a esta zona acuden los pacientes que necesitan atención especial por su estado de salud.

3.2.2 Flujos de Trabajo en la Unidad Quirúrgica del HGUV

En el HGUV se han definido tres flujos de trabajo claramente diferenciados para la unidad quirúrgica, aunque para el presente trabajo se ha decidido concentrarlo en dos ya que se considera que los procesos que incluyen ingreso tienen características muy similares:

- **Cirugía mayor ambulatoria.** En este flujo de trabajo se encuentran todos aquellos pacientes que son sometidos a una cirugía y en el mismo día son dados de alta, sin necesidad de ocupar una cama del hospital. Este tipo de cirugías suelen ser poco invasivas y tienen una recuperación sencilla de tal manera que no es necesario realizar un seguimiento presencial de la evolución del postoperatorio. En el conjunto de datos del que se dispone este tipo de operaciones están identificadas en el campo *Program* como CSI. El flujo de trabajo definido por el hospital para este tipo de operaciones es el que se observa en la ilustración 3.9. El paciente pasa por los distintos estados y cuando pasa un tiempo adecuado de adaptación es dado de alta y puede marcharse.

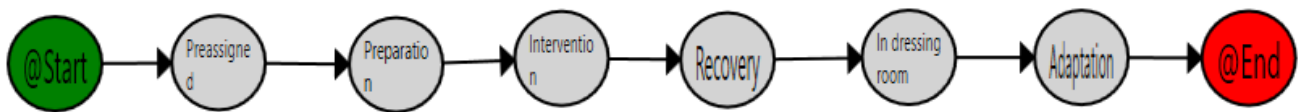


Ilustración 3.9. Flujo de trabajo ideal de un paciente en CSI. Fuente: elaboración propia.

- **Cirugía con ingreso (CCI).** En este flujo de trabajo se encuentran dos tipos de pacientes: aquellos que están hospitalizados, se someten a una cirugía y se quedan hospitalizados; y aquellos que no estaban hospitalizados pero que las características de la operación obligan a que pasen el postoperatorio ingresados. En el conjunto de datos son identificados en el campo *Program* como HOSP o CE, respectivamente. Los flujos de trabajo para estos tipos de operaciones son los correspondientes a las ilustraciones 3.10 y 3.11. Una vez el paciente se recupera de la intervención se debe desasignar la pulsera y el paciente es llevado a su habitación.

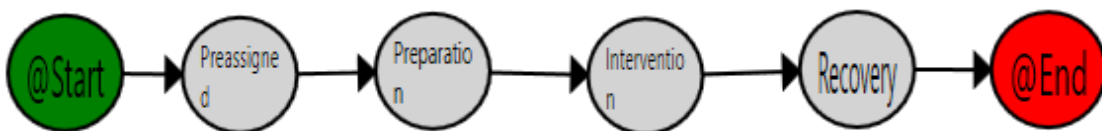


Ilustración 3.10. Flujo de trabajo ideal de un paciente para CE. Fuente: elaboración propia.



Ilustración 3.11. Flujo de trabajo ideal de un paciente para HOSP. Fuente: elaboración propia.

Estos flujos son teóricos y no existe ninguna restricción para evitar que un paciente de cirugía con ingreso pase por una ubicación asociada a la cirugía sin ingreso, por tanto, se pueden encontrar pacientes de CCI cuya traza pasa por adaptación o por el vestuario.

3.3 APLICACIÓN DE MINERÍA DE PROCESOS Y GENERACIÓN DE MODELOS

La herramienta utilizada para realizar la MP y la generación de modelos ha sido una aplicación de minería de procesos (PMAApp) diseñada por Carlos Fernández-Llatas [22], basada en PALIA Suite[7], en la Universitat Politècnica de València. Esta aplicación ha sido personalizada para las necesidades de este trabajo, incluyendo funciones como la creación de modelos y la corrección de errores, además el PMAApp filtra automáticamente aquellas trazas que no contienen un paso por intervención, no teniéndolas en cuenta para el proceso de inferencia.



Ilustración 3.12 Pantalla principal del PMAApp, los cuadrados rojos delimitan zonas y los cuadrados rojos con línea discontinua botones relevantes. Fuente: elaboración propia.

Como puede observarse en la ilustración 3.12, el PMAApp cuenta con 4 zonas diferenciadas:

1. **Zona 1.** Es la zona de representación gráfica de los flujos de trabajo obtenidos de la minería de procesos.
2. **Zona 2.** Es la zona que permite escoger el log a utilizar, además de personalizar cómo va a ser analizado. Como se puede observar aparece un botón superior donde pone *LoadLog* que permite escoger el archivo de extensión csv que se desee, bajo este botón aparece un cuadro de texto editable en el cual se pueden escoger qué nodos se quiere utilizar para la minería escribiendo el nombre del estado y separándolo por comas.
En esta misma zona aparece una casilla de verificación que permite seleccionar si se desea realizar una corrección de errores, función que será explicada más adelante. Bajo esta línea aparece un segundo botón que permite cargar el archivo que se desea utilizar en la corrección de errores.
3. **Zona 3.** Es la zona donde aparecen las distintas herramientas de minería de procesos, como los filtros, la zona que permite cargar mapas de calor, una de visionado de trazas del nodo o transición seleccionadas y otra con estadísticas. Estas funciones no se han utilizado para este trabajo.
4. **Zona 4.** En esta zona aparecen 4 botones. El primero de ellos es el botón de inicio que permite comenzar la minería de datos, el segundo permite entrenar los HMM y el tercero examinarlos. Por otra parte, el cuarto botón permite abrir un diseñador de los archivos que se utilizan para la corrección de errores.

3.3.1 Creación de HMM

En el apartado 2.3.2 se explicó que es posible utilizar el algoritmo *Forward-Backward* para estimar el modelo a partir de un conjunto de observaciones.

El PMAApp utilizado se aprovecha de las capacidades de la MP para descubrir el flujo de trabajo existente y utiliza el algoritmo *Forward-Backward* implementado en PMCode para encontrar las probabilidades a priori. Sin embargo, para poder estimar adecuadamente los modelos es necesario disponer de modelos no cíclicos. Para esto el PMAApp aplica un filtrado que elimina los ciclos especificando la repetición de los modelos en el corpus.

Por otra parte, se entrena un modelo diferente para cada tipo diferente de proceso quirúrgico existente en el corpus, eliminando así la variabilidad asociada al tipo de intervención que se realiza.

3.3.2 Corrección de errores

La corrección de errores es una función del PMAApp que permite filtrar trazas, quedando solamente aquellas se correspondan a un flujo de trabajo o TPA (del inglés *Timed Parallel Automantons*). El usuario de la aplicación puede definir un TPA utilizando el botón *Open Designer*, desde donde se abre editor que permite añadir y borrar transiciones y nodos, como se puede observar en la ilustración 3.13. Una vez generado el TPA podrá ser cargado para

utilizarlo en la corrección de errores y sólo aquellas trazas que se ajusten a él serán utilizadas para la inferencia y para la creación y validación de modelos.

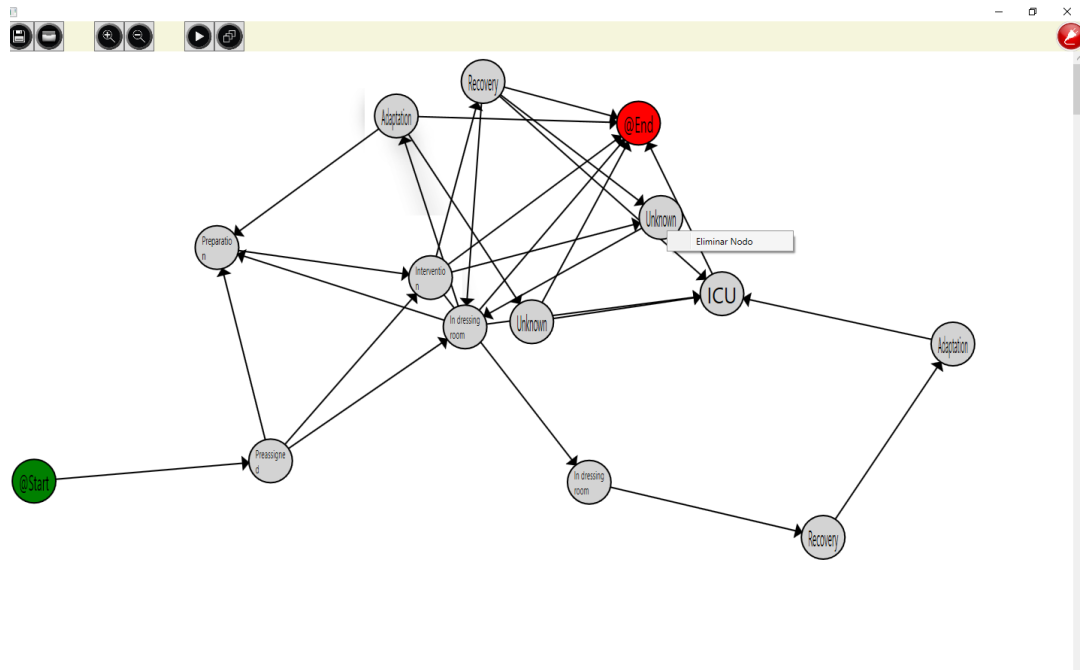


Ilustración 3.13. Ventana de diseño de TPAs del PMApp. Fuente: elaboración propia.

3.4 IDENTIFICACIÓN DE ERRORES EN LOS DATOS

El conocimiento de los datos con los que se va a trabajar es un paso fundamental en cualquier proyecto donde se utilicen técnicas de *Big Data*, por tanto, el primer paso que se realizó fue la identificación de las trazas anómalas presentes en el corpus. Para la correcta identificación de estos errores se tuvieron en cuenta los flujos de trabajo ideales definidos en el apartado 3.2.2.

La realización de un análisis exhaustivo de todos los datos en el corpus supondría una gran cantidad de tiempo, por tanto, se tomó la decisión de analizar 300 trazas distribuidas a lo largo de todo el período del que se dispone y suponer que los errores que se detectasen serían extensibles a la totalidad del corpus.

Los principales errores que se detectaron fueron los siguientes:

- **Reflexiones.** Se trata de trazas en las que el estado del paciente cambia durante un breve período de tiempo y tras el mismo vuelve al nodo en que se encontraba. Este tipo de sucesos puede justificarse de dos formas: que este recorrido sea real debido a alguna alteración en el proceso que está siguiendo el paciente o que, debido a las limitaciones de la tecnología utilizada, una baliza detecte con mayor fuerza la señal de la pulsera cuando en realidad se sigue en la ubicación correspondiente a otra baliza. Tras un tiempo, la baliza correspondiente vuelve a identificar correctamente el tag. Se percibió (y fue confirmado por expertos de la empresa) que la segunda causa era la principal provocadora de estas anomalías. En la ilustración 3.14 se puede observar un ejemplo de este suceso.

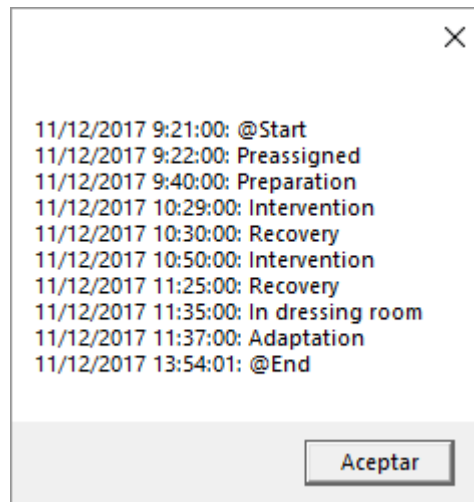


Ilustración 3.14. Ejemplo de traza con reflexión. Fuente: elaboración propia.

- **Desasignación tardía.** Se trata de trazas en las que el tiempo que pasa el paciente en el último nodo es excesivamente largo. Esto sucede si el personal responsable no desasigna la pulsera del paciente una vez este se marcha o si el propio paciente se lleva la pulsera fuera del hospital tras el postoperatorio.
- **Trazas incompletas.** Se trata de trazas que en ningún momento del proceso pasan por intervención. Se puede deber a que, aunque la intervención se encuentre planificada, finalmente no sea realizada, por tanto, el paciente nunca llega a entrar en la zona de quirófano.
Además, algunas de las trazas incompletas se encuentran asociadas a un posible problema con el sistema de asignación de pulseras que provoca múltiples asignaciones de una pulsera a distintos pacientes durante breves períodos de tiempo.

Estos tres tipos de errores que se detectaron tenían en común que provocaban una alteración en la duración de los distintos eventos, lo cual conduce a que los modelos que se generen no se adapten a la realidad de los procesos.

3.5 TÉCNICAS DE MEJORA DE LA CALIDAD DE LA INFORMACIÓN

Ante los errores presentes en el corpus, se implementaron técnicas de detección y corrección de estos utilizando MATLAB R2018b. El objetivo de estas técnicas era conservar el máximo número de trazas para la generación de modelos, además de generar un corpus que se adecuase a la realidad del proceso.

Durante el presente apartado se explicarán las distintas funciones implementadas para la corrección de los errores, el orden seguido en la modificación de los datos y las suposiciones asumidas.

1. **Lectura de los datos.** Se realiza mediante la función *load_data.m*. Esta función lee el archivo que contiene los datos y genera una variable tipo *struct* que almacenará en cada campo otro *struct*, correspondiente a una traza y cuyos campos se corresponden con los campos explicados en el apartado 3.1 exceptuando *Trace* como se puede ver en las ilustraciones 3.15 y 3.16.

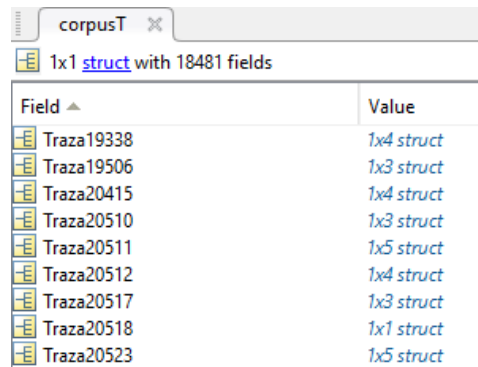


Ilustración 3.15. Aspecto de la salida de *load_data.m*. Fuente: elaboración propia.

Fields	Tag	State	Program	Medical_Service	Surgeon	Surgical_Process	In	Out	Duration
1	'601283000...	'Preassigned'	'CSI'	'OFT'	25191	'FACOEMULSIFICA...	1x1 datetime	1x1 datetime	1x1 duration
2	'601283000...	'Preparation'	'CSI'	'OFT'	25191	'FACOEMULSIFICA...	1x1 datetime	1x1 datetime	1x1 duration
3	'601283000...	'Intervention'	'CSI'	'OFT'	25191	'FACOEMULSIFICA...	1x1 datetime	1x1 datetime	1x1 duration
4	'601283000...	'Recovery'	'CSI'	'OFT'	25191	'FACOEMULSIFICA...	1x1 datetime	1x1 datetime	1x1 duration
5	'601283000...	'In dressing ...'	'CSI'	'OFT'	25191	'FACOEMULSIFICA...	1x1 datetime	1x1 datetime	1x1 duration
6	'601283000...	'Adaptation'	'CSI'	'OFT'	25191	'FACOEMULSIFICA...	1x1 datetime	1x1 datetime	1x1 duration

Ilustración 3.16. Aspecto de una traza correspondiente a un campo de la salida de *load_data.m*. Fuente: elaboración propia.

2. **División de las trazas.** Se realiza mediante la función *div_preas.m*. Esta función recibe la estructura generada en el paso anterior y divide las trazas cada vez que aparece el estado *Preassigned*, proporcionando como salida un corpus similar al de la ilustración 3.12. Se supone que si una traza contiene varias veces el estado *Preassigned* se debe a un error en el sistema de asignación o a un cambio de pulsera antes de comenzar el proceso, por tanto, los eventos que ocurran entre dos estados preasignados no representan la realidad adecuadamente.
3. **Descarte de trazas.** Se realiza mediante la función *erase_traces.m*. Esta función elimina aquellas trazas que no contienen un paso por intervención, puesto que no serán de utilidad a la hora de obtener modelos de predicción de tiempos en el quirófano. Así mismo también elimina aquellas trazas cuyo proceso quirúrgico sea *NULL* puesto que este proceso es el valor que se obtiene en caso de que el campo no fuera rellenado y realmente representa a distintas operaciones, por tanto, no se debe considerar a la hora de generar modelos. Las trazas eliminadas no se introducen en el corpus final pero sí que se mantienen en otro corpus.

4. **Corrección de reflexiones.** Se realiza mediante la función *fix_ref.m*. Esta función convierte en un solo evento los distintos componentes de una reflexión. Para ello se siguen los siguientes pasos:
1. Evalúa si se repite algún estado dentro de la traza, en caso afirmativo, comprueba cuál es y cuantas veces se repite.
 2. Si la reflexión tiene más de un salto, por ejemplo, *Intervention-Recuperation-Intervention-Recuperation- Intervention*, se destina esa traza a un corpus que será cotejado por un experto.
 3. En caso contrario, se comprueba la dirección de la reflexión mediante la relación de duraciones de los eventos y se unen aquellos eventos que realmente formen parte de uno solo. Para la evaluación de la dirección se comprueba si la duración de uno de los componentes del salto es mucho menor (menos del 50%) que el de los otros componentes, ese será el evento que sea absorbido por los otros dos.
 4. Si uno de estos eventos se corresponde con el estado 'Unknown' siempre será el otro evento el que se conserve, puesto que se considera que este estado desconocido se ha debido a una pérdida momentánea de la señal, por tanto, ese tiempo ha debido pasarle en el estado inicial.

En la ilustración 3.14 puede observarse el diagrama de flujo de este algoritmo.

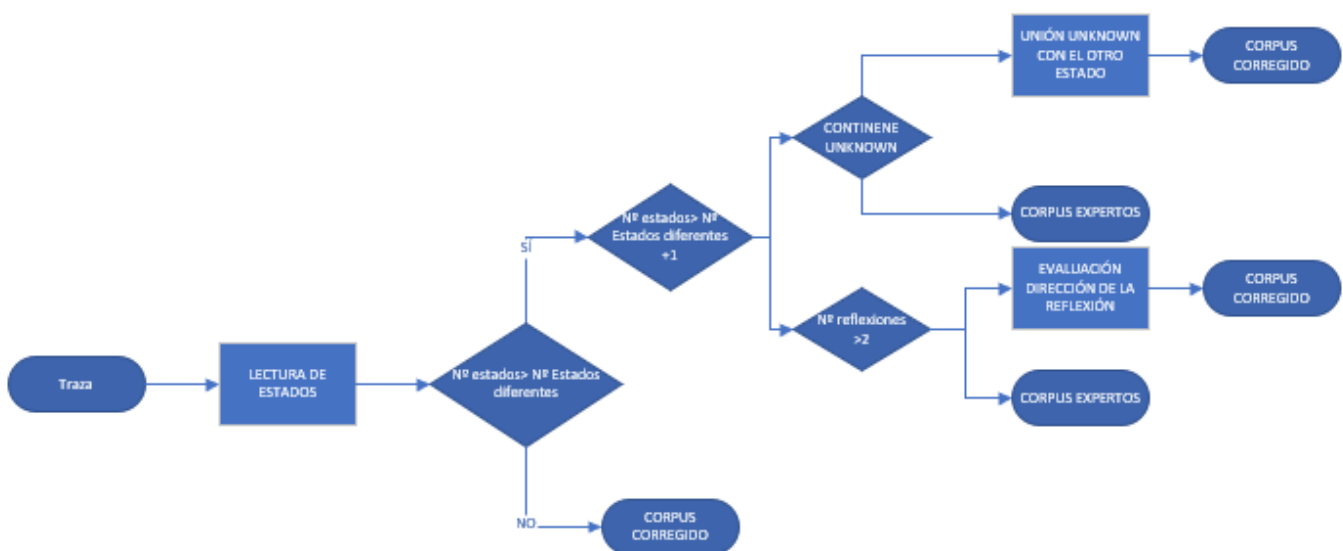


Ilustración 3.17. Diagrama de flujo de la función *fix_ref.m*. Fuente: elaboración propia.

5. **Eliminar Unknowns finales.** Se realiza mediante la función *erase_unknowns.m*. Esta función elimina el último estado de una traza en caso de que este sea *Unknown* puesto que se considera que altera la duración final de la traza sin aportar ninguna información.

6. **Corrección de desasignaciones.** Se realiza mediante la función *fix_desas.m*. Esta función permite realizar una aproximación de la duración que debería tener un evento final que se identifica como desasignación tardía, para ello realiza un estudio de valores anómalos dividiendo por estado y proceso quirúrgico. Esta división se realiza puesto que se considera que la duración de un estado depende del proceso quirúrgico a realizar. Si el valor de un estado en una traza está fuera de 1,5 intervalos intercuartílicos, se sustituirá por la mediana de los estados que no sean considerados anómalos[23]. El uso de la mediana de los estados no anómalos permite obtener una estimación de lo que debería durar ese estado en ese tipo de proceso quirúrgico realmente.

Además, como el servicio quirúrgico cierra a las 22:00, no se permite que ninguna traza supere las 23:00 del día en el que comienza el último estado, ya que de ser así, probablemente se trate de un error. En la ilustración 3.15 puede observarse el diagrama de flujo de esta función.

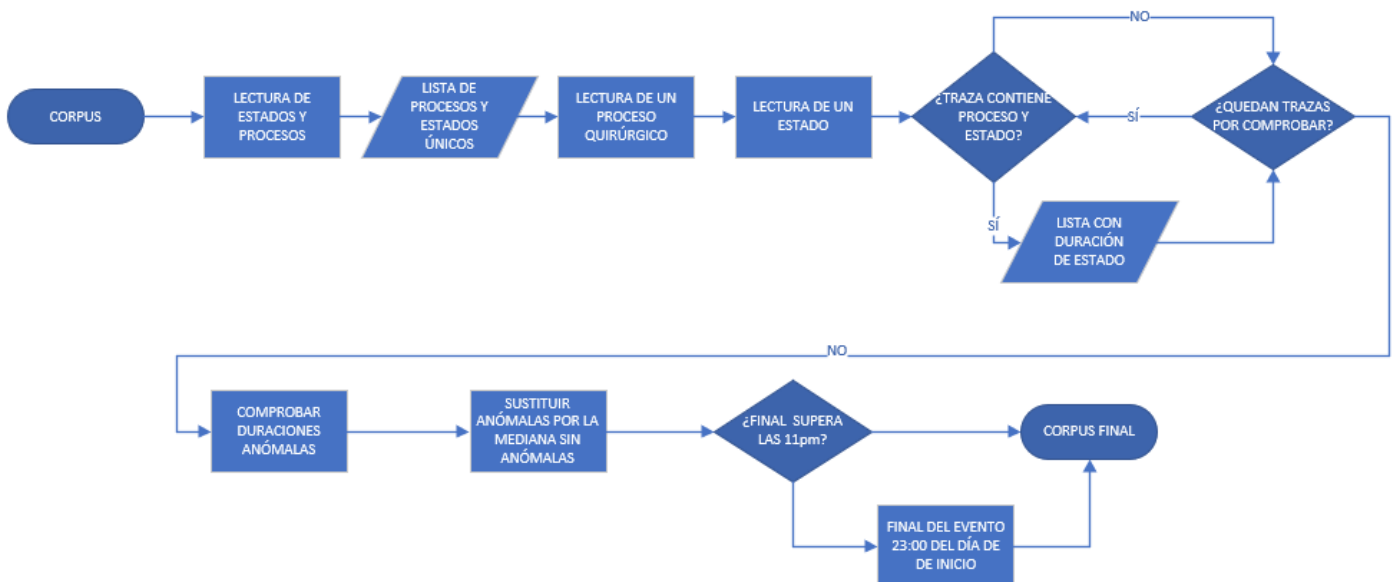


Ilustración 3.18. Diagrama de flujo de la función *fix_desas*. Fuente: elaboración propia.

7. **Escritura de archivo.** Se realiza mediante la función *write_excel.m*. Esta función genera un archivo de Excel que contiene la información con el nuevo corpus con los errores corregidos.

3.6 USO DE LA CORRECCIÓN DE ERRORES DEL PMAPP

Se utilizó la corrección de errores del PMAPP en todos los corpus utilizados, para ello se definieron los flujos de trabajo que se observan en las ilustraciones 3.16 y 3.17. Estos TPAs fueron definidos por el autor en base a lo expuesto en el apartado 3.2.2 permitiendo cierta flexibilidad en el nodo final del evento. Se utilizó la corrección de errores basándose en la suposición de que, si todas las trazas utilizadas para la generación de modelos siguiesen el mismo flujo de trabajo, se generarían modelos más precisos.

Como se puede observar en la ilustración 3.16, el TPA definido para el conjunto de datos al completo es el mismo que para las intervenciones de CSI. Además, se permitió que las trazas pudieran terminar tras el nodo *In dressing room* puesto que se observó que era un suceso muy habitual a pesar de no adaptarse al proceso de la ilustración 3.9.

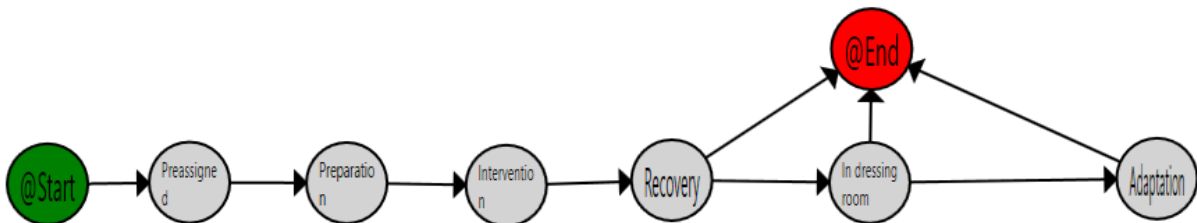


Ilustración 3.19. TPA para corpus original, corpus tras TMC y corpus de CSI. Fuente: elaboración propia

El TPA, correspondiente a la ilustración 3.20 definido para las intervenciones de CCI fue idéntico al de la ilustración 3.10. No se tuvo en cuenta el flujo definido en la ilustración 3.11 al considerarse que el estado *In transit* no aparecía en la mayoría de las trazas de tipo *HOSP*, por tanto, se perderían demasiadas muestras.

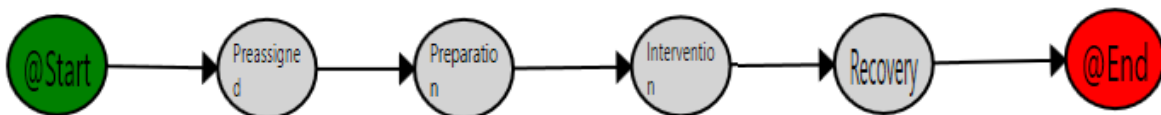


Ilustración 3.20. TPA para corpus de CCI. Fuente: elaboración propia.

3.7 CREACIÓN DE MODELOS DE PREDICCIÓN Y VALIDACIÓN

3.7.1 Entrenamiento de modelos

Se utilizó el PMApp explicado en el apartado 3.3 para la generación de modelos de predicción basados en HMM. Para ello se utilizaron los datos desde el 1 de diciembre de 2017 hasta el 30 de noviembre de 2018 como conjunto de entrenamiento.

Se han realizado modelos utilizando varios corpus distintos, para evaluar la funcionalidad tanto de las TMC (limpio) como del uso de la función de corrección de errores. Así mismo, también se ha realizado una separación del corpus en operaciones de CSI y operaciones de CCI (CE y HOSP) para observar cuáles presentan modelos con mayor fiabilidad una vez aplicadas las TMC.

En la tabla 3.3 se puede observar el número de trazas destinadas al entrenamiento y a validación de cada corpus.

Tabla 3.3. Número de trazas totales, para entrenamiento y para validación en los distintos corpus utilizados. Fuente: elaboración propia.

Corpus	Corrección de errores	Total	Entrenamiento	Validación
Bruto	No	9.255	8.720	535
	Sí	6.917	6.558	359
Limpio	No	8.038	7.598	440
	Sí	7.007	6.607	400
CSI	No	3.786	3.600	186
	Sí	3.618	3.434	184
CCI	No	4.127	3.878	249
	Sí	3.273	3.058	215

3.7.2 Validación de los modelos

Para el proceso de validación se utilizaron los datos desde el 1 de diciembre de 2018 hasta el 28 de diciembre de 2018, siguiéndose el proceso para la validación que aparece en la ilustración 3.21.

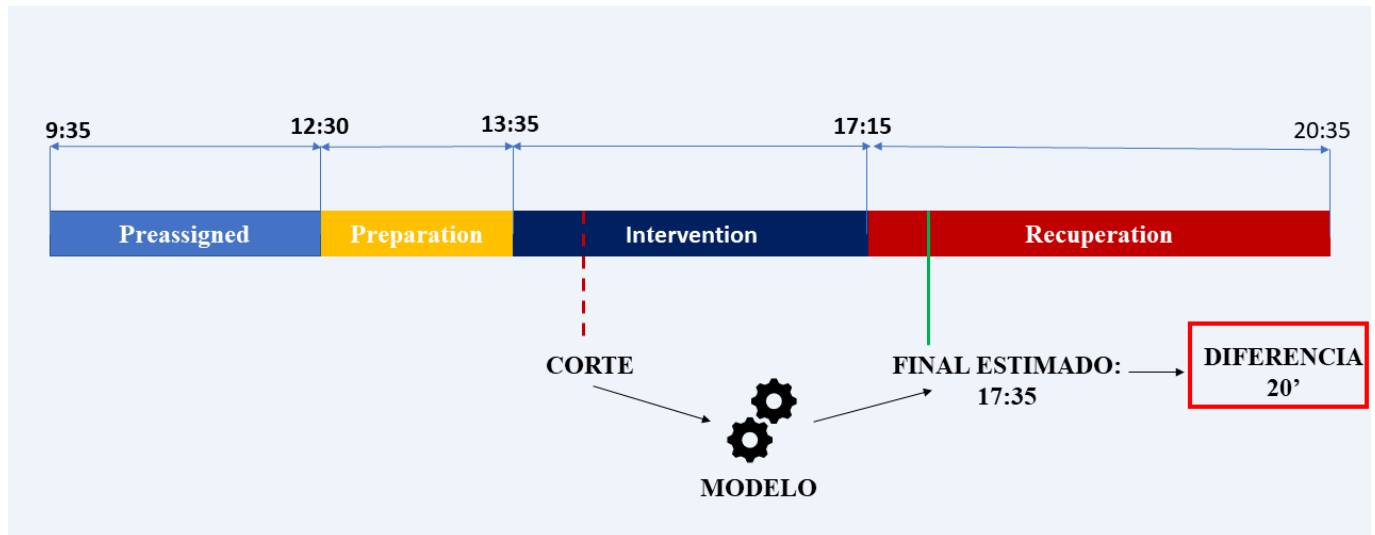


Ilustración 3.21. Funcionamiento del proceso de validación. Fuente: elaboración propia.

Se realizaron cortes horarios en las trazas y se utilizaron los modelos generados para predecir el final del evento en el que se realizó el corte, después se comparó la predicción con el final real del evento y se obtuvo, en minutos, la diferencia entre ambas.

CAPÍTULO 4. RESULTADOS

4.1 MEJORA DE LA CALIDAD DE LOS DATOS

Para evaluar la mejora de la calidad de los datos tras la aplicación de las técnicas desarrolladas, se observó el resultado de la aplicación del algoritmo de MP a los datos mejorados respecto a los datos originales.

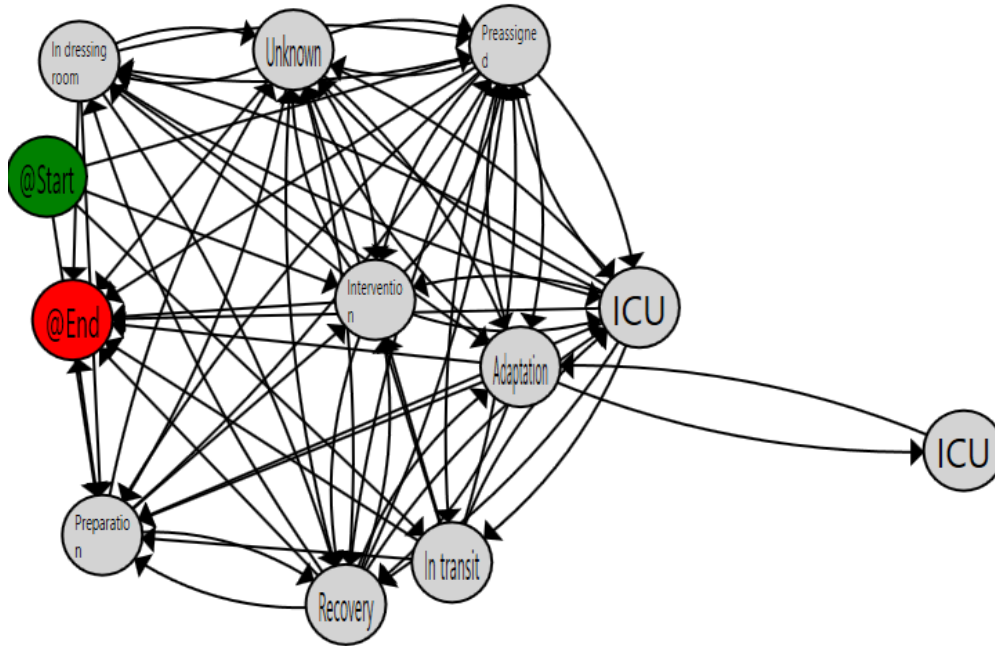


Ilustración 4.1. Resultado de la aplicación del algoritmo de MP al corpus original. Fuente: elaboración propia.

En el primer resultado, obtenido a partir de 9.255 presente en la ilustración 4.1, se pudo observar la existencia de un gran efecto espagueti que hace imposible distinguir el proceso existente en estos datos.

En el resultado tras las TMC, obtenido a partir de 1.217 trazas menos que el anterior, es decir 8.038 y presente en la ilustración 4.2, se siguió observando el efecto espagueti, aunque con menor densidad. Además, se distinguieron nodos que representaban el mismo estado repetidas veces, como *Preparation*, *In dressing room* o *ICU*.

Sin embargo, al analizar las estadísticas de los nodos repetidos, como el número de trazas que atravesaban ese nodo, se observó que eran pocas trazas (números siempre inferiores a la decena) las que hacían que apareciese ese nodo.

Finalmente, es importante destacar que, de las 1.217 trazas perdidas entre ambos corpus, 232 fueron destinadas a un corpus que debería ser evaluado por un experto.

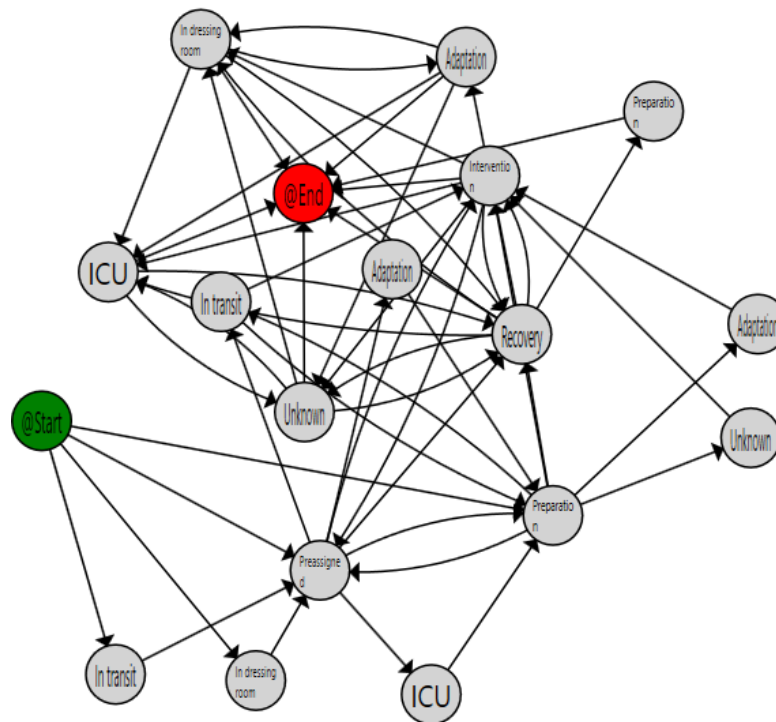


Ilustración 4.2. Resultado de la aplicación del algoritmo de MP al corpus tras las TMC.
Fuente: elaboración propia.

4.2 RESULTADOS DE LA PREDICCIÓN

En el presente apartado se expondrán los resultados de predicción con el conjunto de validación explicado en el apartado 3.7. Debido a la alta variabilidad asociada a los estados *Preassigned*, *In transit*, *ICU* y *Unknown*, a pesar de que se han obtenido predicciones de estos, no se mostrarán en este apartado.

Por otra parte, se decidió que como métrica de evaluación de la mejora de la predicción gracias a las TMC se utilizaría la relación entre el rango intercuartílico de la predicción con los datos originales y la predicción con el corpus corregido. La distribución de las diferencias entre la predicción y los valores reales serán representados mediante diagramas de caja y bigotes no para intentar observar una diferencia significativa sino con la finalidad de observar la variabilidad asociada a la predicción con cada corpus.

4.2.1 Predicción del estado Preparation

La primera predicción realizada con el corpus original para el estado *Preparation* presentó un rango intercuartílico, de 57,04 minutos, es decir, el 50% de las predicciones presentaban una variabilidad de 57 minutos respecto al final real. Al aplicar las técnicas de corrección de errores, el rango no mejoró, como se puede observar en la tabla 4.1.

Tabla 4.1. Predicción del estado *Preparation* con los distintos corpus en minutos. Fuente: elaboración propia.

Corpus	Corrección de errores	Nº de predicciones	Mediana	Rango inter	Relación
Bruto	No	237	-9,57	57,04	1
	Sí	158	-3,81	58,62	1,027
Limpio	No	180	-7,10	39,16	0,69
	Sí	172	-7,45	39,56	0,69
CSI	No	69	-6,91	29,91	0,52
	Sí	69	-7,06	29,90	0,52
CCI	No	107	-9,35	49,80	0,87
	Sí	99	-8,41	46,70	0,82

La aplicación de las TMC sí que supuso una mejora de esta variabilidad, reduciéndola en un 31%, como se puede observar en la ilustración 4.3. Al aplicar la corrección de errores, la predicción no mejoró.

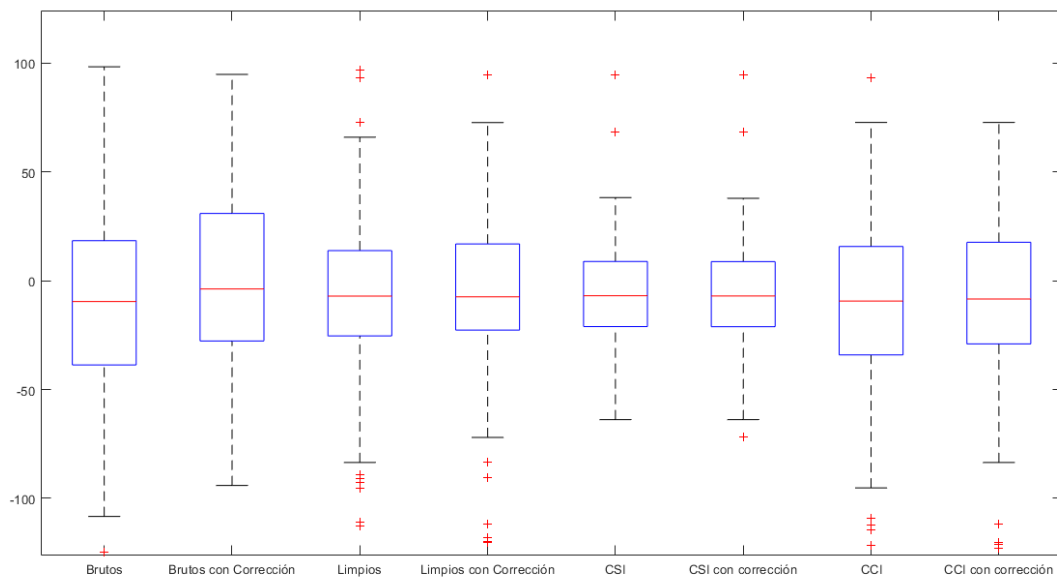


Ilustración 4.3 Diagrama de caja y bigotes con los resultados de la predicción para el estado Preparation, en el eje vertical se representa la diferencia en minutos entre el final predicho y el real. Fuente: elaboración propia.

Finalmente, se pudo observar que la predicción del estado *Preparation* con una menor variabilidad es la asociada con el corpus de CSI, siendo esta en torno a los 30 minutos y no observándose mejora al aplicar las técnicas de corrección de errores. Esta reducción de la variabilidad supuso una reducción superior al 50% respecto a los datos originales.

Mientras tanto, la predicción de la preparación del paciente en CCI presentó una mayor variabilidad (50 minutos sin corrección de errores y 47 con ella) que en CSI, lo que representa una mejora de 13 y 18% respectivamente respecto al corpus original.

4.2.2 Predicción del estado Intervention

En la tabla 4.2 se detalla que la predicción realizada con modelos generados a partir del corpus original presentaba un rango intercuartílico de 133,91 minutos, tras aplicar las técnicas de corrección esta diferencia aumentó en un 18% hasta 158 minutos.

Tabla 4.2. Predicción del estado Intervention con los distintos corpus en minutos. Fuente: elaboración propia.

Corpus	Corrección de errores	Nº de predicciones	Mediana	Rango inter	Relación
Bruto	No	563	19,00	133,91	1
	Sí	347	26,50	158,00	1,18
Limpio	No	449	16,42	119,83	0,89
	Sí	408	22,54	120,58	0,90
CSI	No	89	9,35	42,49	0,32
	Sí	88	9,90	37,36	0,28
CCI	No	346	11,32	135,28	1,01
	Sí	307	26,66	179,03	1,34

Tras las TMC, la variabilidad de la predicción de las intervenciones se redujo en un 11%, quedándose en 120 minutos. El uso de la corrección de errores no mejoró la variabilidad de la predicción, siendo esta de 121 minutos.

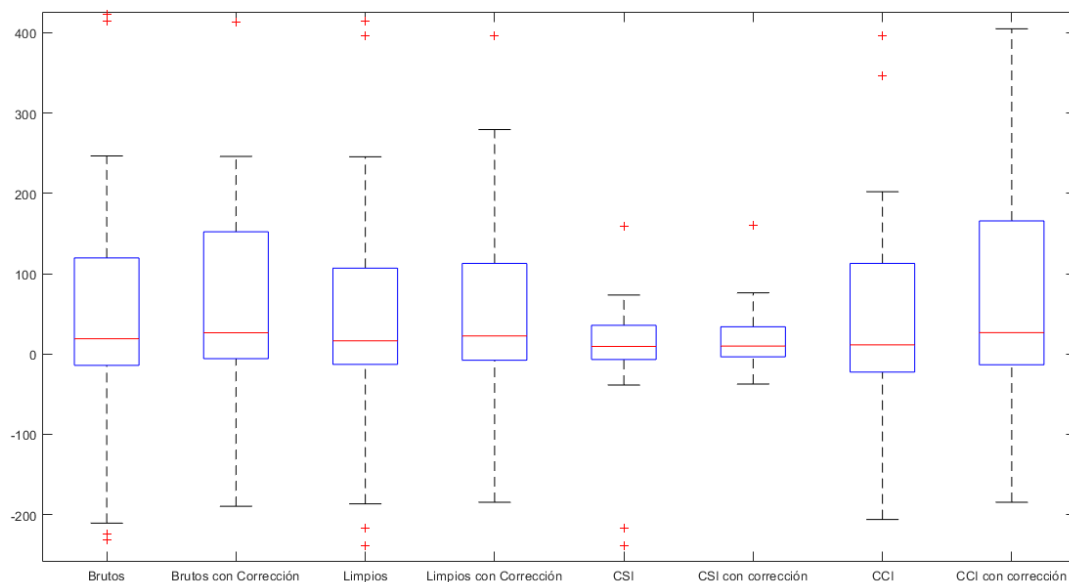


Ilustración 4.4. Diagrama de caja y bigotes con los resultados de la predicción para el estado Intervention, en el eje vertical se representa la diferencia en minutos entre el final predicho y el real. Fuente: elaboración propia.

Al observar la ilustración 4.4 destaca la pequeña variabilidad que presentó la predicción del fin de las intervenciones en CSI, cuyo rango representa una reducción del 68% sin corrección de errores y de 72% con aquellas muestras que se adaptan al TPA definido en el apartado 3.6.

En contraste con esta gran mejora, la predicción de los estados *Intervention* en CCI, experimentó un aumento de la variabilidad de un 1 y un 34 % para los corpus sin corrección y con corrección respectivamente.

4.2.3 Predicción del estado Recovery

En la predicción inicial de este estado, la variabilidad era muy elevada (según la tabla 4.3, 3495,77 minutos), por ello, para una mejor visualización, en la ilustración 4.5 solo se muestra un rango entre 1000 y -1000.

Tabla 4.3. Predicción del estado Recovery con los distintos corpus en minutos. Fuente: elaboración propia.

Corpus	Corrección de errores	Nº de predicciones	Mediana	Rango inter	Relación
Bruto	No	1.435	610,27	3.495,77	1
	Sí	1.148	1.088,88	4.205,49	1.20
Limpio	No	707	94,59	269,53	0,08
	Sí	695	115,90	274,17	0,08
CSI	No	273	234,43	990,56	0,28
	Sí	273	234,47	-786,14	0,28
CCI	No	418	26,94	235,44	0,07
	Sí	408	21,98	202,84	0,06

Como se ha comentado, la variabilidad inicial era muy elevada para la predicción de las recuperaciones, la cual aumentó hasta 4.206 minutos con la corrección de errores, y se vio reducida en un 92% tras aplicar las TMC, quedándose en 270 minutos tanto para el corpus con corrección como el de sin corrección.

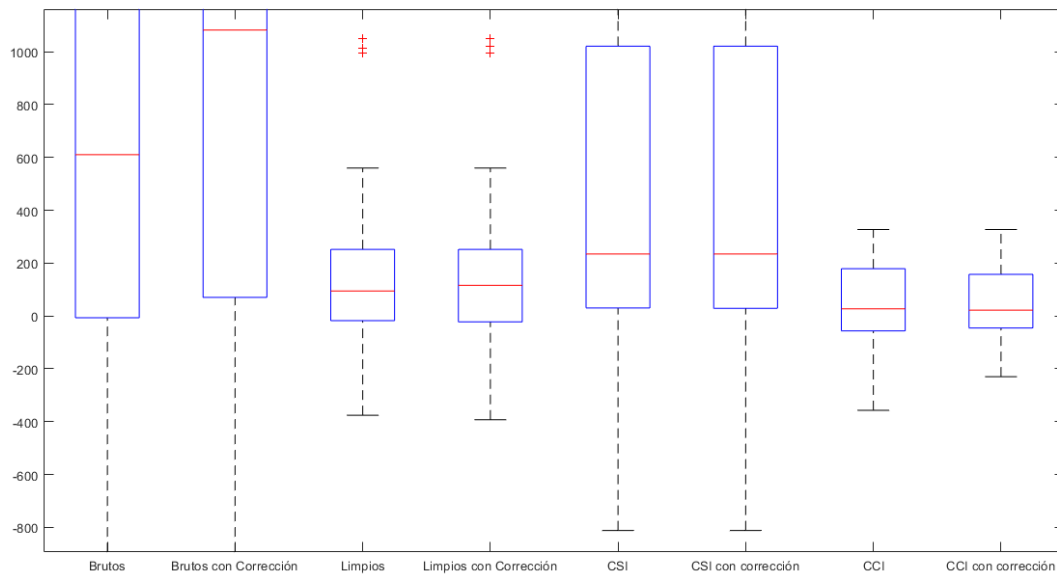


Ilustración 4.5. Diagrama de caja y bigotes con los resultados de la predicción para el estado Recovery, en el eje vertical se representa la diferencia en minutos entre el final predicho y el real. Fuente: elaboración propia.

De los dos tipos de operaciones que se han diferenciado, la que mayor reducción de la variabilidad presentó fue las recuperaciones de CCI, que redujeron en un 93 (sin corrección) y un 94% (con corrección) la variabilidad respecto al corpus original.

Por su parte, las operaciones de CSI presentaron una reducción de un 72% tras las TMC, no observándose mejora tras la aplicación de la corrección de errores.

4.2.4 Predicción del estado *In dressing room*

El estado *In dressing room* no se encuentra presente en el corpus de CCI, por tanto, no se muestra predicción para este corpus en la ilustración 4.6 ni en la tabla 4.4. Al igual que en el apartado anterior se ha limitado el rango de representación para una mejor visualización.

Tabla 4.4. Predicción del estado *In dressing room* con los distintos corpus en minutos.
Fuente: elaboración propia.

Corpus	Corrección de errores	Nº de predicciones	Mediana (min)	Rango inter	Relación
Bruto	No	323	3.987,90	3.189,00	1
	Sí	239	4.048,87	3.855,29	1,21
Limpio	No	62	255,52	302,72	0,10
	Sí	62	255,52	302,72	0,10
CSI	No	62	255,52	302,72	0,10
	Sí	62	255,52	302,72	0,10

Las predicciones iniciales para este estado presentaron una variabilidad de 3.189 minutos, aumentando hasta 3.855 (21%) tras la corrección de errores. Tras la aplicación de las TMC, esta gran variabilidad se vio reducida hasta 302 minutos, lo que representa una mejora de un 90% respecto a la primera predicción. La corrección de errores no presentó ningún efecto en la corrección.

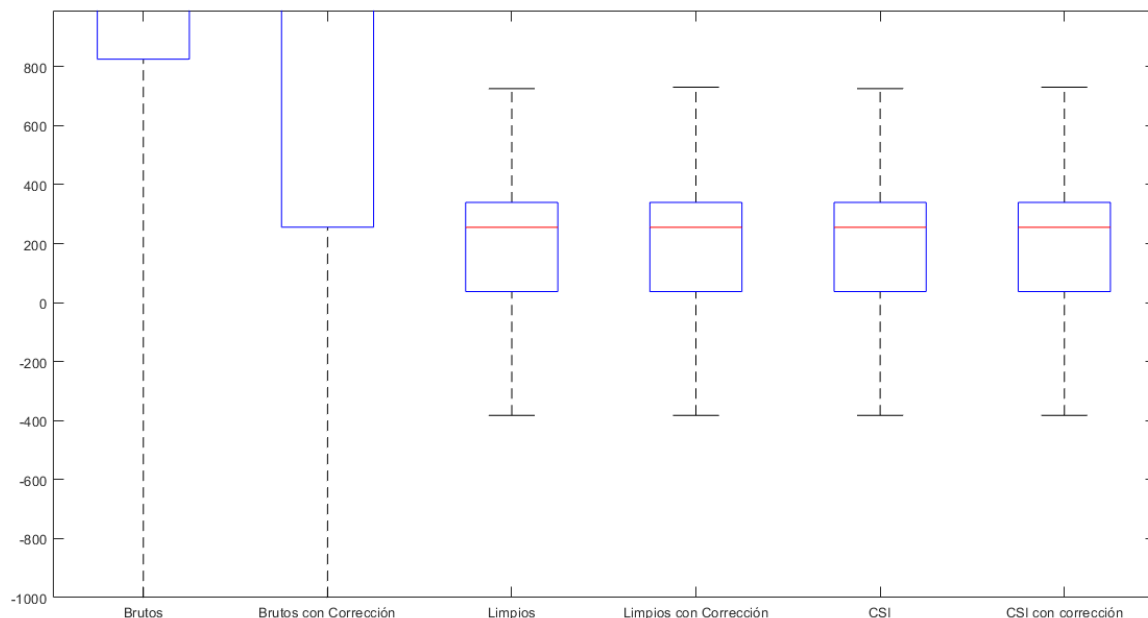


Ilustración 4.6. Diagrama de caja y bigotes con los resultados de la predicción para el estado *In dressing room*, en el eje vertical se representa la diferencia en minutos entre el final predicho y el real. Fuente: elaboración propia.

Las predicciones iniciales para este estado presentaron una variabilidad de 3.189 minutos, aumentando hasta 3.855 (21%) tras la corrección de errores. Tras la aplicación de las TMC, esta gran variabilidad se vio reducida hasta 302 minutos, lo que representa una mejora de un 90% respecto a la primera predicción. La corrección de errores no presentó ningún efecto en la corrección.

Debido a que en el conjunto completo de datos tras las TMC solo las intervenciones de CSI presentaban el estado *In dressing room*, los resultados de la predicción con ambos corpus son iguales.

4.2.5 Predicción del estado Adaptation

El motivo para no presentar en la ilustración 4.7 y la tabla 4.5 las predicciones de CCI es el mismo que se explicó en el apartado anterior. Al igual que para los estados *Recovery* e *In dressing room*, se ha reducido el rango de presentación del diagrama de caja y bigotes para una mejor visualización.

Tabla 4.5 Predicción del estado Adaptation con los distintos corpus en minutos. Fuente: elaboración propia.

Corpus	Corrección de errores	Nº de predicciones	Mediana	Rango interc	Relación
Bruto	No	234	1.045,75	896,04	1
	Sí	145	1.054,47	2290,71	2,56
Limpio	No	42	136,75	227,82	0,25
	Sí	33	82,73	201,03	0,22
CSI	No	42	136,99	228,07	0,25
	Sí	33	82,76	201,03	0,22

La variabilidad de las predicciones iniciales fue de 896 minutos, y tras la corrección de errores esta aumentó hasta 2.291 minutos. Sin embargo, tras las TMC la variabilidad inicial se redujo en un 75% hasta suponer aproximadamente 228,82 minutos.

Desarrollo de Técnicas de Mejora de Calidad de Información en Sistemas de Localización a Tiempo Real para Creación de Modelos de Predicción en Entornos Hospitalarios

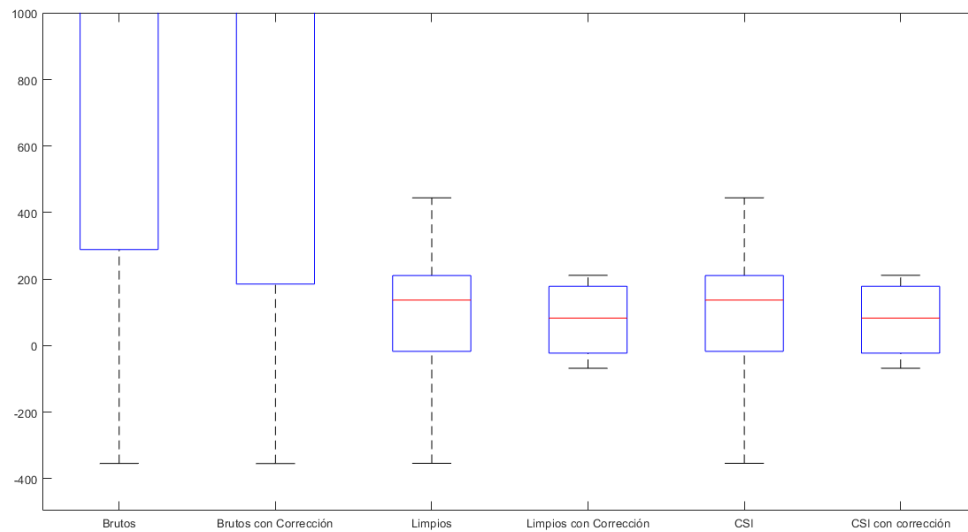


Ilustración 4.7. Diagrama de caja y bigotes con los resultados de la predicción para el estado Adaptation, en el eje vertical se representa la diferencia en minutos entre el final predicho y el real. Fuente: elaboración propia.

La aplicación de la corrección de errores al corpus corregido generó modelos que presentaron una mejora del 78% respecto al corpus original, quedándose en 201 minutos.

Debido a que en el conjunto completo de datos tras las TMC solo las intervenciones de CSI presentaban el estado *Adaptation*, los resultados de la predicción con ambos corpus son iguales.

CAPÍTULO 5. DISCUSIÓN

5.1 EVALUACIÓN DEL EFECTO DE LAS TMC EN LA CALIDAD DE LOS DATOS

El objetivo de este trabajo era desarrollar técnicas de mejora de calidad de la información en RTLS, para así permitir un aprovechamiento de estos dirigido a la creación de modelos de predicción en el bloque quirúrgico.

Un primer aproximamiento en el que se hubieran eliminado todas aquellas trazas que no siguieran los requisitos marcados habría generado una base de datos con muy poco ruido, sin embargo, habría reducido demasiado la cantidad de datos disponible, provocando, probablemente, un sobreentrenamiento de los modelos. Es por esto por lo que el uso de TMC al permitir conservar la mayoría de las trazas, como se observó en el apartado 4.1, debería conseguir modelos más fiables.

En añadido a esto, a pesar de las observaciones realizadas en el 4.1 donde se destacó que habían aparecido nuevos nodos repetidos, estos pertenecían a muy pocas trazas, lo cual se puede interpretar positivamente, puesto que el algoritmo de PALIA separa de esta manera aquellos flujos de trabajo que son estadísticamente diferentes al flujo principal. Esto conlleva que estos flujos diferentes son más fáciles de identificar de lo que lo eran en un principio y solamente son seguidos por una pequeña parte de las trazas, por lo que la calidad de los datos ha mejorado.

Así mismo, si los datos que son destinados al corpus destinado a la evaluación por expertos fueran corregidos de acuerdo con la realidad, estas trazas podrían incluirse en el proceso de entrenamiento, disponiendo de una muestra aún mayor.

5.2 EVALUACIÓN DEL EFECTO DE LAS TMC EN LA PREDICCIÓN

Las TMC desarrolladas tenían como objetivo poder generar modelos de predicción mejores a partir de los datos obtenidos mediante RTLS, lo que se puede identificar como una reducción la variabilidad de la diferencia de minutos entre la predicción y el final real de los eventos.

Esta reducción ha podido observarse en todos los estados detallados durante el apartado 4.2, obteniéndose los mejores resultados en la predicción de los estados *Recovery* e *In dressing room*, con reducciones de la variabilidad de más del 90%. A pesar de esta gran reducción del rango intercuartílico, cabe destacar que la mediana de la desviación en la predicción sigue siendo elevada superando en ambos casos los 200 minutos. Esta gran desviación puede encontrar su explicación en el uso de la función encargada de arreglar las desasignaciones tardías explicada en el apartado 3.3, ya que ambos estados suelen ser el último estado de la traza, y si sufrían una desasignación tardía, se habrá corregido la excesiva duración, pero no se habrá obtenido la duración real, lo cual hace que los modelos no se adapten completamente a la realidad para estos estados. Este mismo suceso se pudo observar en el nodo *Adaptation*, que también suele ser el último estado de la traza, y a pesar de conseguirse una mejora del 78% con el corpus con corrección de errores también muestra una mediana de desviación superior a 1 hora.

Uno de los estados que más interés presentaba a la hora de obtener una predicción era el estado *Intervention*, ya que el quirófano es el factor más limitante para la correcta programación del bloque quirúrgico y poder predecir con anterioridad el momento en el que va a terminar la operación permite preparar con mayor precisión tanto el postoperatorio como al siguiente paciente que va a ser intervenido a continuación. Se observó que, en general, se obtuvo una mejora del 10% de la variabilidad tras las TMC con una mediana de desviación de 16 minutos; pero la mejora más impactante fue la producida en las intervenciones de CSI que presentan una mejor hasta de un 72% tras aplicar las TMC y la corrección de errores. Es por tanto un tipo de intervención que cuanto más se ajuste al proceso predefinido, mejor predicción se podrá realizar de su duración.

En contraste con este hecho, destaca que las TMC no han mejorado la predicción en las intervenciones del corpus de CCI, lo cual muestra que la complejidad de este tipo de operaciones les añade un componente de variabilidad que no se ve corregido con las TMC.

Finalmente cabe destacar que los corpus generados tras el uso de la corrección de errores del PMAApp no han mejorado la predicción de manera llamativa e incluso la han empeorado. Esto puede deberse a que los TPAs definidos no eliminan suficiente ruido en los corpus.

CAPÍTULO 6. CONCLUSIONES

El objetivo principal de este trabajo era desarrollar TMC de la información que permitieran aprovechar los datos generados por RTLS para la generación de modelos de predicción intrahospitalarios. Durante todo el trabajo, se han identificado aquellos eventos que producían un empeoramiento de la calidad de la información y se han elaborado algoritmos capaces de detectarlos y corregirlos gracias al conocimiento adquirido sobre el funcionamiento de las tecnologías RTLS y sobre el bloque quirúrgico del HGUV.

Por esto, el autor considera que los RTLS ofrecen una oportunidad única de obtener información sobre los procesos a los que se somete un paciente cuando entra en el bloque quirúrgico, además de mejorar la calidad de la atención a los familiares. Esta información generada puede ser usada para la elaboración de modelos de predicción, lo cual permitirá la automatización de la programación de la actividad de los quirófanos permitiendo una orquestación del flujo de pacientes que conduzca a ofrecer una atención sanitaria mejor, con menos listas de esperas y reduciendo los costes del servicio.

Sin embargo, mientras estas soluciones no sean aplicadas de manera adecuada por los usuarios y no se mejore el algoritmo de localización hasta conseguir una información totalmente fiable, será necesario utilizar técnicas de detección de errores y mejora de la calidad de esta información para que los modelos lleguen a poder ser utilizados con fiabilidad.

En el presente trabajo se ha demostrado que la variabilidad en la predicción mejora de manera notable tras el uso de las TMC, por tanto, son una solución viable para mejorar la optimización del bloque quirúrgico mediante RTLS.

6.1 LÍNEAS FUTURAS

Tras la realización del trabajo, se ha considerado que existen múltiples líneas futuras que permitirían mejorar el desarrollo realizado, así como los resultados obtenidos. Estas mejoras son:

- El uso de un corpus de datos con más pacientes para la generación de modelos, teniéndose así modelos más fiables de cada tipo de operación.
- El uso de un corpus cuyos nodos se correspondan con la localización en lugar de con el estado permitiría una corrección de las reflexiones más precisa.
- La implementación de las TMC en una plataforma como el PMApp permitiría una integración, generando un producto capaz de corregir los datos y generar los modelos.
- La integración de la plataforma del punto anterior con el sistema de información del RTLS permitiría su utilización para optimizar tanto el bloque quirúrgico como otro tipo de procesos hospitalarios.

- El uso de los corpus y modelos obtenidos para el desarrollo del proyecto *Operating Room 4.0* comprendido en el marco *SME Instrument* del programa europeo Horizon 2020.

Así mismo, y debido al trabajo realizado, la empresa MYSPHERA se ha puesto en contacto con el autor para que continúe con su trabajo de análisis de los datos por los RTLS una vez haya completado los estudios, debido a la utilidad que les aporta conocer el funcionamiento de los sistemas instalados en cada hospital y disponer de nuevas herramientas que generen un valor añadido a su producto.

CAPÍTULO 7. REFERENCIAS

- [1] “Ministerio de Sanidad, Consumo y Bienestar Social - Portal Estadístico del SNS - Sanidad en datos.” [Online]. Available: <https://www.msbs.gob.es/estadEstudios/sanidadDatos/home.htm>. [Accessed: 12-Jun-2019].
- [2] jlozano, “Gasto Sanitario Informe Anual del Sistema Nacional de Salud 2017 INFORMES, ESTUDIOS E INVESTIGACIÓN 2018 MINISTERIO DE SANIDAD, SERVICIOS SOCIALES E IGUALDAD.”
- [3] N. Martin, “Using Indoor Location System Data to Enhance the Quality of Healthcare Event Logs: Opportunities and Challenges,” 2019, pp. 226–238.
- [4] J. Vissers and R. Beech, *Health operations management : patient flow logistics in health care*. Routledge, 2005.
- [5] P. Aparicio Azcárraga, M. Alfaro Latorre Dirección Coordinación, M. De Bustos Guadaño Amparo Montesinos Alonso, D. García Pancorbo Inmaculada Moreno Faraco, and V. Valls Chamizo Enrique Magdaleno Muñoz María José Rodríguez López, “Sistema de Información de Atención Especializada(SIAE): Indicadores hospitalarios. Evolución 2010-2016.” 2017.
- [6] “Listas de espera en la Sanidad pública, últimos datos oficiales.” [Online]. Available: <https://www.epdata.es/datos/listas-espera-sanidad-publica/24/espana/106>. [Accessed: 13-Jun-2019].
- [7] C. Fernandez-Llatas, A. Lizondo, E. Monton, J. M. Benedi, and V. Traver, “Process mining methodology for health process tracking using real-time indoor location systems,” *Sensors (Switzerland)*, vol. 15, no. 12, pp. 29821–29840, 2015.
- [8] “MYSOPHERA, IoT & RTLS for Healthcare.” [Online]. Available: <https://www.mysophera.com/?lang=es>. [Accessed: 27-May-2019].
- [9] W. M. P. van der Aalst *et al.*, *Lecture Notes in Business Information Processing - IEEE taskforce on process mining*. 2010.
- [10] E. Rojas, J. Munoz-Gama, M. Sepúlveda, and D. Capurro, “Process mining in healthcare: A literature review,” *J. Biomed. Inform.*, vol. 61, pp. 224–236, Jun. 2016.
- [11] C. Fernández-Llatas, J. M. Benedi, J. M. García-Gómez, and V. Traver, “Process mining for individualized behavior modeling using wireless tracking in nursing homes,” *Sensors (Switzerland)*, vol. 13, no. 11, pp. 15434–15451, 2013.
- [12] C. Fernandez-Llatas, A. Martinez-Millana, A. Martinez-Romero, J. M. Benedi, and V. Traver, “Diabetes care related process modelling using Process Mining techniques. Lessons learned in the application of Interactive Pattern Recognition: coping with the Spaghetti Effect,” in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 2127–2130.
- [13] P. S. Stepaniak, C. Heij, and G. De Vries, “Modeling and prediction of surgical procedure times,” *Stat. Neerl.*, vol. 64, no. 1, pp. 1–18, 2010.
- [14] E. Choi *et al.*, “RETAIN: An Interpretable Predictive Model for Healthcare using Reverse Time Attention Mechanism.”

- [15] M. Abir, M. M. Davis, P. Sankar, A. C. Wong, and S. C. Wang, “Design of a Model to Predict Surge Capacity Bottlenecks for Burn Mass Casualties at a Large Academic Medical Center,” *Prehosp. Disaster Med.*, vol. 28, no. 1, pp. 23–32, Feb. 2013.
- [16] L. R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
- [17] C. de Souza, “Hidden Markov Models in C#.” [Online]. Available: https://www.codeproject.com/Articles/69647/Hidden-Markov-Models-in-C#_articleTop. [Accessed: 19-May-2019].
- [18] S. E. Levinson, “Hidden Markov model speech recognition arrangement,” *The Journal of the Acoustical Society of America*, vol. 86, no. 6, pp. 2478–2478, 2005.
- [19] “CHGUV - Inicio.” [Online]. Available: <http://chguv.san.gva.es/inicio>. [Accessed: 17-Jun-2019].
- [20] Hospital General Universitari and Generalitat Valenciana, “ACTIVIDAD ASISTENCIAL 2018. RESUMEN DEPARTAMENTO VALENCIA DEPARTAMENTO VALENCIA DEPARTAMENTO VALENCIA DEPARTAMENTO VALENCIA--HOSPITAL GENERAL HOSPITAL GENERAL HOSPITAL GENERAL HOSPITAL GENERAL,” 2019.
- [21] “El Hospital General de Valencia realiza más de 120 intervenciones para implantar válvulas aórticas y mitrales - Alicante News.” [Online]. Available: <https://alicantenews.es/tendencias/el-hospital-general-de-valencia-realiza-mas-de-120-intervenciones-para-implantar-valvulas-aorticas-y-mitrales/>. [Accessed: 17-Jun-2019].
- [22] G. Ibanez-Sanchez *et al.*, “Toward Value-Based Healthcare through Interactive Process Mining in Emergency Rooms: The Stroke Case,” *Int. J. Environ. Res. Public Health*, vol. 16, no. 10, p. 1783, May 2019.
- [23] D. Cousineau, “OutliersDetectionAndTreatment.pdf,” vol. 3, no. 1, pp. 58–67, 2011.

PRESUPUESTO

Índice del Presupuesto

CAPÍTULO 1.	INTRODUCCIÓN	1
CAPÍTULO 2.	PRESPUESTO DESGLOSADO	2
2.1	Coste de Mano de Obra	2
2.2	Coste de Maquinaria y Software	3
2.3	Coste de Precios Unitarios.....	4
2.4	Cuadro de Precios Descompuestos.....	5
2.5	Presupuestos Parciales	10
2.6	Presupuesto de Ejecución por Contrata	12

CAPÍTULO 1. INTRODUCCIÓN

El objetivo de este presupuesto es el cálculo de los costes asociados con la realización de este proyecto. Para ello se ha dividido el proyecto en las siguientes fases:

1. Definición del proyecto y obtención de los datos.
2. Análisis del funcionamiento de los sistemas de localización a tiempo real (RTLS) y el bloque quirúrgico del Hospital General Universitario de València (HGUV).
3. Análisis de los datos en busca de problemas de calidad.
4. Desarrollo de técnicas de mejora de la calidad (TMC).
5. Validación de las TMC mediante la generación de modelos de predicción.
6. Redacción de la memoria y defensa del proyecto.

Todos los costes asociados a la mano de obra han tenido en cuenta las tasas y la seguridad social, así como los precios de materiales no incluían el IVA.

CAPÍTULO 2. PRESUPUESTO DESGLOSADO

2.1 COSTE DE MANO DE OBRA

CUADRO DE PRECIOS DE MANO DE OBRA						
Nº Recurso	Código	Descripción del Recurso	Unidad	Precio Unitario (€)	Cantidad	Total (€)
1	MO01	Investigador (Tutor)	€/hora	25	10	250
2	MO02	Investigador (Cotutor 1)	€/hora	25	50	1.250
3	MO03	Investigador (Cotutor 2)	€/hora	25	5	125
4	MO04	Ingeniero de Telecomunicaciones	€/hora	25	10	250
5	MO05	Ingeniero Biomédico Junior	€/hora	12,5	311	3.887,5
					Total	5.762

2.2 COSTE DE MAQUINARIA Y SOFTWARE

CUADRO DE PRECIOS DE MAQUINARIA Y SOFTWARE							
Nº Recurso	Código	Descripción del Recurso	Unidad	Precio Unitario (€)	Cantidad	Factor de Amort.	Total (€)
1	MAQ01	Portátil Acer 15,6" Aspire E5-571G Intel Core i5 5200U	€	495	1	6/48	61,88
2	MAQ02	Portátil DELL Inspiron 155000	€	5.000	1	6/18	1.666,67
3	MAQ03	Microsoft Office 365	€/licencia	79	1	6/48	9,88
4	MAQ04	Visual Studio Enterprise	€/licencia	2727	1	6/12	1.363,5
5	MAQ05	MATLAB R2018b	€/licencia	250	1	6/12	125
						Total	3.351,93

2.3 COSTE DE PRECIOS UNITARIOS

CUADRO DE PRECIOS UNITARIOS			
Nº Actividad	Código	Descripción de las unidades de obra	Precio (€)
1	UO1	Definición del proyecto y obtención de los datos	2.016,05
2	UO2	Análisis del funcionamiento de RTLS y el bloque quirúrgico del HGUV	275
3	UO3	Análisis de los datos en busca de problemas de calidad	262,5
4	UO4	Desarrollo de TMC	2.562,5
5	UO5	Validación de las TMC mediante la generación de modelos de predicción	2.851
6	UO6	Redacción de la memoria y defensa del proyecto	712,5

2.4 CUADRO DE PRECIOS DESCOMPUESTOS

CUADRO DE PRECIOS DESCOMPUESTOS						
1	DEFINICIÓN DEL PROYECTO Y OBTENCIÓN DE LOS DATOS					
Nº	Capítulo					
1.1	Reunión inicial					
	Código recurso	Descripción del recurso	Precio unitario (€)	Unidad	Cantidad	Precio (€)
	MO01	Investigador (Tutor)	25	h	1	25
	MO02	Investigador (Cotutor 1)	25	h	1	25
	MO03	Investigador (Cotutor 2)	25	h	1	25
	MO05	Ingeniero Biomédico Junior	12,5	h	1	12,5
	MAQ01	Portátil Acer 15,6" Aspire E5-571G Intel Core i5 5200U	61,88	u	1	61,88
	MAQ02	Portátil DELL Inspiron 155000	1.666,67	u	1	1.666,67
					Subtotal	1.816,05
1.2	Traspaso de datos entre MYSPHERA y el grupo investigador					
	MO01	Investigador (Cotutor 1)	25	h	1	25
	MO04	Ingeniero de Telecomunicaciones	25	h	4	100

Desarrollo de Técnicas de Mejora de Calidad de Información en Sistemas de Localización a Tiempo Real para Creación de Modelos de Predicción en Entornos Hospitalarios

	MO05	Ingeniero Biomédico Junior	12,5	h	4	50
					Subtotal	200
					Total	2.016,05
2.	ANÁLISIS DEL FUNCIONAMIENTO DE RTLS Y BLOQUE QUIRÚRGICO DEL HGUV					
2.1	Reunión con experto de MYSPHERA					
	MO04	Ingeniero de Telecomunicaciones	25	h	4	100
	MO05	Ingeniero Biomédico Junior	12,5	h	4	50
					Subtotal	150
	Estudio de literatura por parte del autor					
	MO05	Ingeniero Biomédico Junior	12,5	h	10	125
					Subtotal	125
					Total	275
3.	ANÁLISIS DE LOS DATOS EN BUSCA DE PROBLEMAS DE CALIDAD					
3.1	Caracterización del corpus de datos					
	MO01	Investigador (Cotutor 1)	25	h	1	25
	MO05	Ingeniero Biomédico Junior	12,5	h	1	12,5
					Subtotal	37,5
3.2	Búsqueda de problemas de calidad y verificación					
	MO01	Investigador	25	h	2	50

Desarrollo de Técnicas de Mejora de Calidad de Información en Sistemas de Localización a Tiempo Real para Creación de Modelos de Predicción en Entornos Hospitalarios

		(Cotutor 1)				
	MO04	Ingeniero de Telecomunicaciones	25	h	2	100
	MO05	Ingeniero Biomédico Junior	12,5	h	6	75
					Subtotal	225
					Total	262,5
4	DESARROLLO DE TMC					
4.1	Diseño de los algoritmos					
	MO01	Investigador (Cotutor 1)	25	h	5	125
	MO05	Ingeniero Biomédico Junior	12,5	h	15	187,5
					Subtotal	312,5
4.2	Implementación de los algoritmos					
	MO01	Investigador (Cotutor 1)	25	h	10	250
	MO05	Ingeniero Biomédico Junior	12,5	h	150	1.875
	MAQ05	MATLAB R2018b	125	u	1	125
					Subtotal	2.250
					Total	2.562,5
5	VALIDACIÓN DE LAS TMC MEDIANTE LA GENERACIÓN DE MODELOS DE PREDICCIÓN					
5.1	Diseño y personalización del PMApp					
	MO01	Investigador	25	h	24	600

Desarrollo de Técnicas de Mejora de Calidad de Información en Sistemas de Localización a Tiempo Real para Creación de Modelos de Predicción en Entornos Hospitalarios

		(Cotutor 1)				
	MO05	Ingeniero Biomédico Junior	12,5	h	6	75
	MAQ04	Visual Studio Enterprise	1363,5	u	1	1.363,5
					Subtotal	2.038,5
5.2	Diseño del corpus para la generación de modelos					
	MO05	Ingeniero Biomédico Junior	12,5	h	1	12,5
					Subtotal	12,5
5.3	Generación y validación de modelos de predicción					
	MO05	Ingeniero Biomédico Junior	12,5	h	55	687,5
					Subtotal	687,5
5.4	Evaluación de los resultados					
	MO02	Investigador (Cotutor 1)	25	h	2	50
	MO05	Ingeniero Biomédico Junior	12,5	h	5	62,5
					Subtotal	112,5
					Total	2.851
6	REDACCIÓN DE LA MEMORIA Y DEFENSA DEL PROYECTO					
6.1	Redacción de la memoria					
	MO01	Investigador (Tutor)	25	h	5	125
	MO05	Ingeniero Biomédico Junior	12,5	h	48	225
	MAQ03	Microsoft Office 365	9,88	u	1	9,88
					Subtotal	359,88

Desarrollo de Técnicas de Mejora de Calidad de Información en Sistemas de Localización a Tiempo Real para Creación de Modelos de Predicción en Entornos Hospitalarios

6.2	Corrección de la memoria					
	MO01	Investigador (Tutor)	25	h	3	75
	MO02	Investigador (Cotutor 1)	25	h	3	75
	MO03	Investigador (Cotutor 2)	25	h	3	75
	MO05	Ingeniero Biomédico Junior	12,5	h	2	25
					Subtotal	250
6.3	Preparación de la defensa					
	MO01	Investigador (Tutor)	25	h	1	25
	MO02	Investigador (Cotutor 1)	25	h	1	25
	MO03	Investigador (Cotutor 2)	25	h	1	25
	MO05	Ingeniero Biomédico Junior	12,5	h	3	37,5
					Subtotal	112,5
					Total	722,38
			TOTAL PRECIOS DESCOMPUESTOS			8.779,43

2.5 PRESUPUESTOS PARCIALES

CUADRO DE PRESUPUESTOS PARCIALES		
1.	DEFINICIÓN DEL PROYECTO Y OBTENCIÓN DE LOS DATOS	
Nº	Capítulo	Importe (€)
1.1	Reunión inicial	1.816,05
1.2	Traspasso de datos entre MYSPHERA y el grupo investigador	200
	Subtotal	2016,05
2.	ANÁLISIS DEL FUNCIONAMIENTO DE RTLS Y BLOQUE QUIRÚRGICO DEL HGUV	
2.1	Reunión con experto de MYSPHERA	150
2.2	Estudio de literatura por parte del autor	125
	Subtotal	275
3.	ANÁLISIS DE LOS DATOS EN BUSCA DE PROBLEMAS DE CALIDAD	
3.1	Caracterización del corpus de datos	37,5
3.2	Búsqueda de problemas de calidad y verificación	225
	Subtotal	262,5
4	DESARROLLO DE TMC	
4.1	Diseño de los algoritmos	312,5
4.2	Implementación de los algoritmos	2.250
	Subtotal	2.562,5
5	VALIDACIÓN DE LAS TMC MEDIANTE LA GENERACIÓN DE MODELOS DE PREDICCIÓN	
5.1	Diseño y personalización del PMAApp	2.038,5

Desarrollo de Técnicas de Mejora de Calidad de Información en Sistemas de Localización a Tiempo Real para Creación de Modelos de Predicción en Entornos Hospitalarios

5.2	Diseño del corpus para la generación de modelos	12,5
5.3	Generación y validación de modelos de predicción	687,5
5.4	Evaluación de los resultados	112,5
	Subtotal	2.851
6	REDACCIÓN DE LA MEMORIA Y DEFENSA DEL PROYECTO	
6.1	Redacción de la memoria	350
6.2	Corrección de la memoria	250
6.3	Preparación de la defensa	112,5
	Subtotal	712,5
TOTAL PRESUPUESTOS PARCIALES		8.779,43

2.6 PRESUPUESTO DE EJECUCIÓN POR CONTRATA

Para la realización del presupuesto de ejecución por contrata hay que considerar el presupuesto de ejecución material calculado, los gastos generales, el beneficio industrial y el IVA.

Los gastos generales, es decir, los gastos indirectos, se calculan como el 13% del presupuesto de ejecución material, fijado de forma general según el artículo 131, apartado 1.a) del Reglamento General de la Ley de Contratos de las Administraciones Públicas, aprobado por Real Decreto 1098/2001, de 12 de octubre. El beneficio industrial, será el 6% del presupuesto de ejecución material, correspondiente al beneficio económico esperado del proyecto.

El IVA se corresponde con un 21% del presupuesto de ejecución material, añadiendo los gastos generales y el beneficio industrial.

PRESUPUESTO TOTAL DE EJECUCIÓN POR CONTRATA		
Código	Descripción de las unidades de obra	Coste (€)
UO1	Definición del proyecto y obtención de los datos	2.016,05
UO2	Análisis del funcionamiento de RTLS y el bloque quirúrgico del HGUV	275
UO3	Análisis de los datos en busca de problemas de calidad	262,5
UO4	Desarrollo de TMC	2.562,5
UO5	Validación de las TMC mediante la generación de modelos de predicción	2.851
UO6	Redacción de la memoria y defensa del proyecto	712,5
Presupuesto de Ejecución Material		8.779,43
Gastos generales (13%)		1.141,33
Beneficio industrial (6%)		526,77
Presupuesto de Ejecución por Contrata		10.447,53
IVA (21%)		2.193,98
Presupuesto Global de Licitación		12.641,51

El presupuesto global de licitación asciende a la cantidad de **doce mil seiscientos cuarenta y uno con cincuenta y un euros**.