

Analítica de datos en procesos asistenciales

Debido a que este Trabajo de Fin de Máster fue realizado en colaboración con Gradient, quien posee la propiedad industrial de los desarrollos realizados durante el proyecto. Se considera conveniente no publicar su contenido debido a términos de confidencialidad, por lo que se opta elaborar este resumen que de forma general explica el trabajo realizado.

Durante este trabajo se detallan las técnicas aplicadas a dos problemas prácticos del proceso de urgencias para dos hospitales:

- La predicción diaria del número de hospitalizados
- La detección de tiempos de espera atípicos entre dos actividades del proceso de urgencias.

Estas tareas buscan en esencia ayudar al usuario diariamente o con frecuencia para el manejo de cuellos de botella o de la preparación futura de recursos.

Predicción del número de hospitalizados

Con el análisis y predicción de la cantidad de pacientes hospitalizados, buscamos concretamente: responder al área de urgencias cuáles son las covariables exteriores que influyen en el número de hospitalizados y generar un modelo predictivo capaz de aproximar nuestra variable de interés. Para ello, no solo nos planteamos el uso de variables recopiladas del propio proceso de urgencias, al igual consideraremos variables meteorológicas que históricamente demuestran efectos en la tendencia de ingresos hospitalarios.

Una de las formas en las que podemos representar una variable con múltiples registros a lo largo del tiempo puede ser planteada como un vector o una variable funcional. Como solución al problema de predicción del número de hospitalizados se presentan los modelos aditivos funcionales con respuesta escalar, en específico, la utilización del algoritmo de selección de variables (**fregre.gsam.vs**¹), el cual nos proporciona la mejor combinación de covariables que más información aportan para la predicción del número de hospitalizaciones. Adicional a esto, nuestra aplicación considera la dependencia temporal de nuestros datos, indicando el retardo que produce el mínimo error cuadrático medio entre todos los modelos evaluados.

Analizando los resultados obtenidos son bastante prometedores, ya que aun contando con limitada información, fuimos capaces proporcionar modelos que ofrecían resultados interesantes sobre la relación entre el número de hospitalizaciones y las covariables presentadas. Podemos destacar que, entre los análisis realizados, obtuvimos una ligera mejora en los resultados al descartar las variables meteorológicas del algoritmo de selección.

Detección de tiempos de espera atípicos

Enfocados en la optimización del proceso y mejor flujo del área de urgencias, se espera que la detección en tiempo real de atípicos se presente como un módulo intuitivo y visual capaz de informar al usuario de urgencias sobre cuáles pacientes en el proceso presentan tiempos de espera muy elevados o anómalos entre dos actividades.

¹ Detalles sobre la función: `fregre.gsam.vs`
<https://www.rdocumentation.org/packages/fda.usc/versions/2.0.2/topics/fregre.gsam.vs>

Para la detección de atípicos consideraremos diferentes variables que van acorde a las características del paciente (clasificación de triaje) y condiciones propias del área de urgencias (tiempo de espera, hora de inicio en la actividad y número de pacientes activos).

Como solución a esta tarea presentamos aplicaciones del aprendizaje automático enfocadas para este tipo de escenarios: Isolation forest (IF), Local outlier factor (LOF) y Robust random cut forest (RRCF) . Detectamos cómo se comportan tales metodologías mediante un estudio de simulación, obteniendo para cada método una idea de la certeza que nos pueda presentar sus resultados. Imitando nuestro caso real, generamos por simulación muestras e introducimos anomalías en ellas. De acuerdo a los resultados obtenidos, en cada caso se observó que el método IF presentó una alta sensibilidad clasificando una mayor cantidad de puntos como anomalías, pero sin embargo es el modelo que presenta menor certeza en sus resultados. En la metodología LOF se percibe una menor sensibilidad a la hora de determinar si un dato es atípico o no y sus resultados son intermedios en cuanto a la certeza. Por último, el método RRCF es el método que mayor consistencia presenta en sus resultados, obteniendo para cada uno de los casos simulados una alta certeza en su clasificación.

En nuestro caso práctico se espera que los resultados sean similares a los obtenidos en nuestro estudio simulado y efectivamente se perciben algunas de esas características observadas en la aplicación de cada método. De los modelos presentados podríamos decir que el modelo RRCF es el más adecuado para este estudio en particular, ya que presenta una alta certeza y consistencia, con la ventaja de que es capaz de utilizarse en una base de datos en tiempo real.

Conclusiones

Las metodologías presentadas se caracterizan por ser flexibles, siendo capaces de mejorar sus resultados al introducir variables adicionales. En general, se tiene que cada una de las aplicaciones presentan soluciones de actualidad acordes a las finalidades planteadas para cada una de las tareas del proyecto.