



ITESO, Universidad
Jesuita de Guadalajara

Maestría de Ciencia de Datos

Investigación, Desarrollo e Innovación III

Notas sobre el artículo *Support vector
regression for warranty claim
forecasting*

Estudiante: Daniel Nuño

Profesor: Dr. Jaime Emanuel AlcaláS Temores

Fecha entrega: 14 de octubre, 2022

El artículo

Link: https://kar.kent.ac.uk/31997/1/EJOR_UoK.pdf

Premisa y objetivos

Pronosticar el número de reclamos de garantía es vital para los fabricantes/proveedores de garantía en la preparación de planes fiscales. Enfatiza en la literatura existente, la cual han desarrollado una serie de técnicas de pronóstico, incluidos modelos de Poisson logarítmicos lineales, filtro de Kalman, modelos de series temporales y modelos de redes neuronales artificiales como el perceptron multicapa.

Es decir, el objetivo es la cantidad de reclamos y la pregunta de interés es el pronóstico a meses futuros para una correcta planeación fiscal.

El artículo establece tres suposiciones clave:

1. Los reclamos de garantía reportados en los últimos meses pueden ser más importantes en la previsión de futuros reclamos de garantía que los informados en los meses anteriores.
2. Las tasas de reparación (es decir, el número total de reclamaciones dividido por el número total de productos en servicio), puede provocar la pérdida de información a través de dicha operación de aritmética.
3. Los datos obtenidos como reclamos de garantías están divididos en reclamos válidos, faltan reclamos no hechos (FBNR), o reclamos inválidos (RBNF).
4. La póliza de garantía del producto es a largo plazo, y caduca cuando se ha anunciado la obsolescencia tecnológica.

Datos

Los datos son reclamos agregados mensualmente, donde i ($i=1,2,\dots,m$) en la primera columna representa el mes de la venta. La pregunta de interés es el pronóstico de los reclamos $x_{i,n_0+1}, \dots, x_{i,n_0+k}$ para $(n_0 + 1) - th, (n_0 + 2) - th, \dots, (n_0 + k) - th$ meses dados los datos de los reclamos, $x_{i,n_0+1}, \dots, x_{i,n_0+k}$, hasta el $(n_0) - th$ meses disponibles.

La explicación de la siguiente tabla es para un producto. Comenzó a venderse en mes 1, han transcurridos n meses desde que comenzó a venderse y se ha vendido durante m meses $n \geq m$. Los datos en *meses en servicio* son los reclamos de garantías de forma cuantitativo pueden ser los totales o un ratio de cantidad de reclamos por cada x de unidades vendidas. No es una matriz cuadrada. Y la verdad es que después la normaliza para que cada fila sean los meses en servicio y la cantidad de reclamos sea el valor esperado de reclamos.

Warranty claims ($n \geq m$)

Sold month	Sold amount	Months in service (MIS)							
		1	2	...	$n - m + 1$	$n - m + 2$...	$n - 1$	n
1	N_1	x_{11}	x_{12}	...	$x_{1,n-m+1}$	$x_{1,n-m+2}$...	$x_{1,n-1}$	$x_{1,n}$
2	N_2	x_{21}	x_{22}	...	$x_{2,n-m+1}$	$x_{2,n-m+2}$...	$x_{2,n-1}$	
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots		
$m - 1$	N_{m-1}	$x_{m-1,1}$	$x_{m-1,2}$...	$x_{m-1,n-m+1}$	$x_{m-1,n-m+2}$			
m	N_m	$x_{m,1}$	$x_{m,2}$...	$x_{m,n-m+1}$				

La elección del modelo apropiado para pronosticar el número esperado de reclamos depende de una variedad de factores tales como: la naturaleza del producto (reparable o no reparable), el alcance de la cobertura de la garantía (total o limitada), la política de sustitución (prorrataada o gratuita) y la dimensión de la garantía (unidimensional o bidimensional), etc.

Los autores consideran usar el número de reclamos en lugar de el ratio de fallas por productos vendidos para evitar la pérdida de información, ya que qué productos (1) pueden fallar y no ser reportados (FBNR) o (2) reportados pero sin falla (RBNF).

Modelo

El modelo se enfoca en la estrategia de reparación, si es reparable o no reparable y utiliza dos funciones de probabilidad:

Productos reparables dentro de un intervalo (t_1, t_2) , el número esperado de reclamos de garantía para los productos vendidos en el i -th mes es dado por:

$$N_{ri}(t_1, t_2) = N_i \int_{t_1}^{t_2} (1 - p(v))\lambda(v)dv + N_i \int_{t_1}^{t_2} q(v)dv$$

En esta ecuación el primer término representa el número de reclamos hechos en cierta porción de productos dictada por la probabilidad de que se levante un reclamo por un producto fallido $1 - p(t)$. El segundo termino representa el numero de reclamos hechos en productos que no fallaron. Eso es, algunos reclamos son improcedentes para la reparación

Productos no reparables se les entrega una igual en lugar de reparar el defectuoso.

$$N_{nri}(t_1, t_2) = N_i \left(\int_0^{t_2} (1 + \mu(t_2 - v))f(v)(1 - p(v))dv - \int_0^{t_1} (1 + \mu(t_1 - v))f(v)(1 - p(v))dv \right) + N_i \int_{t_1}^{t_2} q(v)dv$$

Como en la ecuación anterior, el primer termino es el número de reclamos hechos en algunos productos defectuosos dependiendo en $1 - p(t)$, y el segundo termino representa el número de reclamos hechos de los productos que no han fallado.

Funciones f , p y q se pueden estimar con ayuda de un experto o con datos históricos.

El acercamiento sugerido es un modelo de regresión de vector de soporte ponderado con datos de reclamos originales como variables de entrada y salida del modelo, y un modelo de serie temporal basado en SVR para el caso en que solo se disponga de datos de tasa de reclamos.

Método SVR contra otros

Utiliza dos conjuntos de datos. Para cada conjunto separa en entrenamiento y validación. Para ambos casos la validación es de 6 meses. El primer conjunto tiene los meses en servicio, el ratio de reparaciones por cada 1000 carros del modelo actual y el ratio de reparaciones por cada 1000 carros del modelo anterior.

Time in service	Prior model year	Current model year	Time in service	Prior model year	Current model year	Time in service	Prior model year	Current model year
1	0.55	1.27	13	78.2	53.3	25	167	172
2	4.68	1.58	14	84.0	61.8	26	173	186
3	9.88	1.58	15	84.7	75.9	27	179	199
4	12.8	7.71	16	89.0	90.8	28	185	211
5	15.5	11.4	17	97.8	102	29	187	228
6	20.0	18.7	18	110	105	30	189	249
7	25.9	26.3	19	125	113	31	196	271
8	33.1	35.4	20	139	125	32	205	290
9	37.5	42.0	20	139	125	33	211	300
10	45.4	46.4	22	158	139	34	213	328
11	56.4	48.2	23	163	159	35	215	375
12	68.9	48.8	24	166	165	36	221	430

El segundo conjunto solo contiene los meses en servicio y h que es un ratio de reclamos por cada 10000 productos.

MIS	h	MIS	h	MIS	h	MIS	h	MIS	h
1	3.89	8	21.08	15	20.20	22	17.84	29	14.43
2	9.42	9	22.69	16	19.98	23	17.46	30	19.12
3	14.23	10	20.23	17	20.05	24	14.70	31	17.92
4	17.41	11	23.07	18	16.68	25	14.68	32	23.39
5	20.71	12	21.94	19	17.36	26	18.29	33	17.28
6	21.87	13	19.79	20	16.56	27	18.51	34	41.91
7	22.61	14	18.25	21	18.58	28	19.71		

Plantea dos modelos a validar:

- De serie de tiempo para cuando se tiene los ratios de reclamo como variable dependiente.
- De regresión cuando se tiene los reclamos originales como variable dependiente.

Los modelos y parámetros están optimizados para obtener el MSE más pequeño para la serie de tiempo y el wSSE más pequeño para la regresión.

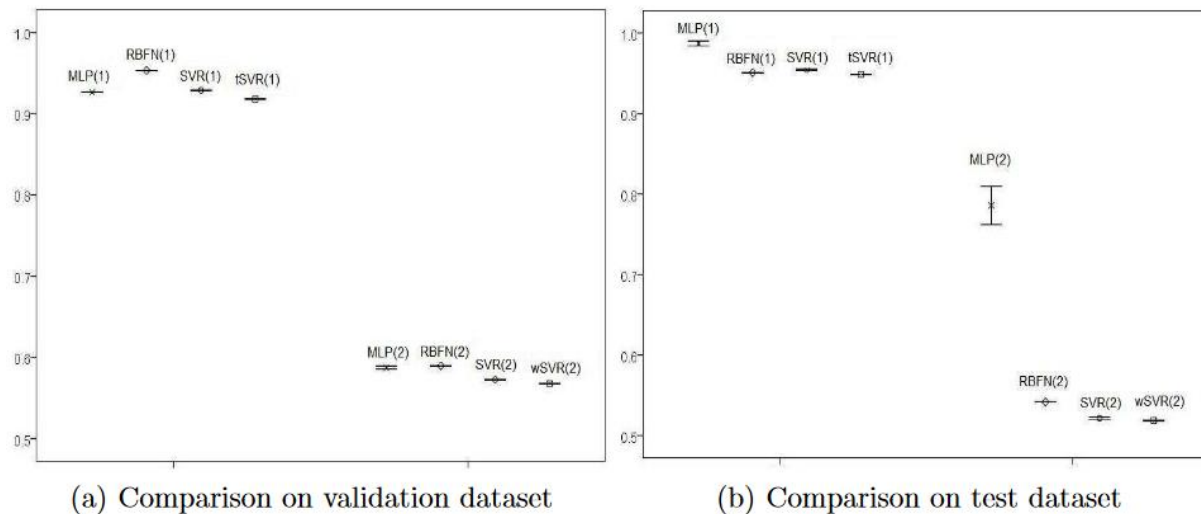


Fig. 1. Comparison between forecasting models 1 and 2. (The labels ended with (1) are time series models, and those labels ended with (2) are regression models.)

Conclusión

Este artículo utiliza 5 páginas para explicar matemáticamente cómo funcionan los reclamos y un modelado que no vuelve a usar y que es difícil de entender. Sin esas páginas hubiera sido más fácil de entender todo. Al final solo toma los datos normalizados de otros artículos, que representa los meses en servicio y la ratio de reclamos y les aplica un Weighted Support Vector Machines. Nunca aplica una regresión lineal, ingeniería de características o promedios móviles/ARIMA que a lo mejor funcionaba mejor.

En mi opinión y como lo entendí, es un paper que podría ser un ejercicio de Kaggle.

Su hallazgo más importante es que el rendimiento de predicción de los modelos con reclamaciones de garantía originales como variables dependientes del modelo supera al de aquellos modelos con ratio de reclamaciones como variables dependientes del modelo. Es decir, el modelo de regresión es mejor que el de las series de tiempo.

Parece no tomar en cuenta que usualmente las garantías tienen un periodo de gracia, 6,12, 18 meses, etc. porque nunca lo nota explícitamente y el valor de m y los meses en servicio no está acotado.

Los autores hacen énfasis en la literatura, pero no necesariamente en los métodos utilizados en la industria. Es un artículo técnico que desarrolla el uso de unos dos algoritmos con buenos resultados, pero no aborda la posible implementación completa para un fabricante o proveedor de garantía.

Tampoco hace referencia del componente que falló, principalmente en productos electrónicos con componentes fácilmente dañables o de un lote malo. Hay otras implementaciones que hacen uso de telemetría de los productos para saber la probabilidad de que un producto falle, aunque el periodo útil del pronóstico es más corto y no es usado para planeación fiscal sino para una mejor atención al cliente.

Diferencias con mi proyecto de aplicación profesional

Mi proyecto de aplicación profesional es la estimación de costos operativos de la organización de Customer Support que se encarga de todos los reclamos de garantías. Existen tres componentes principales:

- reclamos de garantía a través de llamadas. Costos variable o semifijo de call center.
- si el reclamo es valido la reparación o intercambio se realiza. Costos variables de cadena de suministro e ingenieros en función de las reparaciones, es decir, mientras más reparaciones más costo total.
- costos fijos de empleados administrativos, soporte y directivos, además de inversiones e infraestructura de IT.

La mayor parte de los costos y más difíciles de estimar es el componente de la reparación, por lo que es muy importante tener una metodología como la del artículo para saber cuántas reparaciones se tendrán que hacer en el futuro y multiplicarlo por un factor monetario adecuado y así tener el costo.

Aunque, el artículo utiliza el termino *reclamos* no vagamente. La metodología, precedentes, supuestos y datos parece indicar que la intención es estimar la cantidad de llamadas que se esperan recibir ya que el modelo y ecuaciones consideran productos que fallaron, pero no fueron reportados (FBNR) y productos que no fallaron y fueron reportados (RBNF, reclamos improcedentes).

La implementación del artículo y los datos parece aplicar a productos por separado, en mi caso podría ser un modelo en particular de computadora o impresora, por ejemplo. Actualmente no cuento con los datos de meses en servicio y desconozco si sea fácil obtenerlos. Además, HP cuenta con miles de productos y con diferentes periodos de garantía, después de ese periodo no proceden. Al contrario de lo planteado en el artículo, los meses en servicio podrían estar limitados a 12 meses.

Posibles aplicaciones

Me quedo con la implementación weighted Support Vector Machines y la implementación a series de tiempo, ya que me pueden ser útiles.