Temporary Series for Structured Decision Making of Train Line 1

Nicomedes Toledo Ito, Kenny Dany Toledo Calla and Yordan Nicolas Toledo Calla

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

July 16, 2018
TEMPORARY SERIES FOR STRUCTURED DECISION MAKING OF TRAIN LINE 1
LAS SERIES TEMPORALES PARA LA TOMA DE DECISIONES PROGRAMADAS DE LA LÍNEA 1 DEL TREN.

Toledo Ito, Nicomedes, Mg¹, Toledo Calla, Kenny Dany², Ing-M.Sc.²-Toledo Calla, Yordan Nicolas.³, student³
UNIVERSIDAD ANDINA NESTOR CACERES VELASQUEZ, sollata@hotmail.com
UNIVERSIDAD NACIONAL DE INGENIERIA,ktoledoc@uni.pe²
UNIVERSIDAD ANDINA NESTOR CACERES VELASQUEZ, toledo.vpr@hotmail.com³

Resumen— En esta investigación se puede apreciar que las Series Temporales se comportan mejor al darnos Coeficiente de Determinación Pearson ($R^2$) de 0.98 realizando predicciones en estados No lineales de la demanda de pasajeros, que las predicciones realizadas en otros programas lineales como el Excel que nos entrega un Coeficiente de Determinación Pearson ($R^2$) de 0.557, esta ventaja nos permite tomar mejores [6] Decisiones programadas de dotar con mayores trenes a las tres estaciones del Tren de la Línea 1 con mayor correlación de demanda respecto a las demás estaciones de: 0.94 San Borja (SBS), 0.972 Cabitos (CAB) y 0.956 Grau demanda, para evitar pérdidas futuras por desconocer cuáles son las estaciones que más influyen respecto al total, con resultados de 258,33,274.67 y 447 respectivamente lo cual significa que se debe prestar mayor atención a la estación Grau y finalmente a Cabitos para mejorar el Servicio de atención oportuna al cliente.

Palabras Claves—Series Temporales, Toma de decisiones Programadas y no Programadas, Regresión Ridge.

Summary— In this investigation it can be seen that the Temporary Series behave better by giving us Pearson's Coefficient of Determination ($R^2$) of 0.98 by making predictions in non-linear states of passenger demand, than the predictions made in other linear programs such as Excel that we it delivers a Pearson’s Determination Coefficient ($R^2$) of 0.557, this advantage allows us to take better [6] Decisions programmed to provide with greater trains the three stations of the Line 1 Train with the highest correlation of demand with respect to the other stations of: 0.94 San Borja, 0.972 Cabitos and 0.956 Grau demand, to avoid future losses due to not knowing which are the most influential stations with respect to the total, with results of 258,33,274.67 and 447 respectively, which means that more attention must be paid to the Grau station and finally to Cabitos to improve the service of timely attention to the client.

Key Words—Temporary Series, Decision Making Programmed and Unscheduled, Ridge Regression.

I. INTRODUCTION (INTRODUCCION)

Los métodos de Serie de Tiempo en la predicción de tráfico comenzó por Ahmed y Cook, enfocados en predicciones de modelos Autoregresivos Integrado Movimiento en Promedio (ARIMA). [1], gracias a este aporte podemos generar mayores opciones para resolver diferentes situaciones de decisión de la vida en la alta gerencia [2].

A. Entrenamiento de Ecuaciones (Training)

Hoy en día requerimos disminuir nuestras incertidumbres de la vida cuando deseamos tomar alguna decisión también nos hacemos una pregunta: ¿Por qué nos da esto tanto trabajo? Es sencillo, porque no sabemos tomar bien las decisiones. Debo determinar cuáles son las estaciones del Tren con mayor correlación? La única manera de aumentar realmente sus posibilidades de tomar una buena decisión es aprender a aplicar un buen procedimiento para ese fin: el que le dé el mejor resultado con una mínima pérdida de tiempo, energía, dinero y compostura [3]. Así mismo debe estar apoyado en un cálculo numérico que nos ayude a mejorar nuestro punto de vista, a continuación se hizo un cálculo preliminar solo con tres estaciones entre Cabitos, Angamos y Nicolás Arriola del total por la correlación que presenta:

$$y = \beta_0 + \beta_1 CAB + \beta_2 ANG + \beta_3 NAR$$

$$y = 5928.76 + 8448CAB + 4.575ANG + 14605NAR$$

Sin embargo la que mejor se comporta con una buena correlación es entre las Estaciones de: 1.32 San Borja, 1.33 Cabitos y 0.96 Grau y así mismo el error es menos respecto a la anterior ecuación.

Fig. 1: Plano clave, Línea 1 del Tren Eléctrico de Lima.
Fuente: Línea 1 del Tren Eléctrico de Lima.

II. DESARROLLO DE LA INVESTIGACIÓN.

A. Figuras

En la figura 2 nos muestra el mejor comportamiento con un error estándar residual de 0.00092, Figure 2 muestra la ecuación lineal de predicción.

![Figura 2](imagen2.png)

Fig. 2 Total de pasajeros que viajan según los datos de la tabla 1

![Figura 3](imagen3.png)

Figura 3 se aprecia que la linealidad no tiene la capacidad de comportarse no linealmente según el Programa Excel, y es menos confiable por darnos un Coeficiente de Determinación de 0.5577.

B. METODOS DE CONTRACCION Y GRAFICOS.

RIDGE REGRESSION: Se realiza la penalización de

$$\hat{\beta}^\text{ridge} = \beta \left( \sum_{i=1}^{N} \left( y_i - \beta_0 - \sum_{j=1}^{p} x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right)$$

METODO LASSO: Elimina las variables que no aportan

$$\hat{\beta}^\text{lasso} = \beta \left( \frac{1}{2} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right)$$

En la figura cuatro analizamos todas las estaciones para quedarnos con las que tienen buena correlación.
La Figura 04: Se aprecia cuáles son las variables que inciden más para la optimización de los datos de ingreso. Se ilustra en el eje “x”, el log. de Lambda y el eje “y” los β.

II. PREDICCIÓN

Con estos datos de la Tabla 1, procedemos a realizar la predicción en un horizonte futuro, en la Fig. 04 los resultados son comparados entre la Demanda real con la predicción por lo tanto en un futuro cercano se podrá tomar mejores decisiones de planeamiento, con la implementación de estas tecnologías para el control por medio del monitoreo satelital.

Fuente: Línea 1 del Tren Eléctrico de Lima.

En la Figura 05, se puede visualizar gráficamente la verdadera demanda generada por los datos de aforos de la Tabla 1 de las estaciones analizadas, luego el gráfico de color azul es la representación de la predicción realizada por el modelo matemático planteado.

III. VALOR DE LOS ERRORES Y CORRELACIÓN

> pred=predict(modelo,NDatos, interval="confidence")
> RMSE(pred,NDatos1$s1$)

[1] 1.941305e-05

Fechas diarias: Lunes, Martes, Miércoles, Jueves, Viernes
Demanda por estaciones: VES, PIN, PUM, VMA, MAU, SJU, ATO, JCH, AYA, CAB, ANG, SBS, CULT, NAR, GAM, GRAU, ELA, PRES, CAA, PIR, JAR, POS, SCA, SMA, SRO, BAY, BAY

Tabla 1: Datos diarios, empleados desde 04-05-2012 al 11-26-14

En la Figura 05, se puede visualizar gráficamente la verdadera demanda generada por los datos de aforos de la Tabla 1 de las estaciones analizadas, luego el gráfico de color azul es la representación de la predicción realizada por el modelo matemático planteado.
En esta última columna se aprecia, que la mayor correlación se encuentra en la estación Naranjal (Nar) con un valor de 0.978.

| Estación | Coeficiente | Error Estimado | Valor t | Pr(>|t|) |
|----------|-------------|----------------|---------|---------|
| Intercept| 9.428e-03   | 5.348e-02      | 1.760e-01 | 0.860 |
| VES      | 1.00e+00    | 1.305e-06      | 7.662e+05 | <2e-16 |
| PIN      | 1.00e+00    | 3.342e-06      | 3.131e+05 | <2e-16 |
| PUM      | 1.00e+00    | 3.194e-06      | 3.131e+05 | <2e-16 |
| VMA      | 1.00e+00    | 1.589e-06      | 1.892e+05 | <2e-16 |
| MAU      | 1.00e+00    | 1.847e-06      | 1.892e+05 | <2e-16 |
| SJU      | 1.00e+00    | 2.557e-06      | 3.911e+05 | <2e-16 |
| ATO      | 1.00e+00    | 1.787e-06      | 3.911e+05 | <2e-16 |
| JCH      | 1.00e+00    | 4.379e-06      | 2.225e+05 | <2e-16 |
| AYA      | 1.00e+00    | 5.285e-06      | 1.892e+05 | <2e-16 |
| CAB      | 1.00e+00    | 3.194e-06      | 3.911e+05 | <2e-16 |
| ANG      | 1.00e+00    | 1.341e-06      | 1.892e+05 | <2e-16 |
| SBS      | 1.00e+00    | 3.765e-06      | 3.911e+05 | <2e-16 |
| CULT     | 1.00e+00    | 9.135e-07      | 9.135e-07 | 0.393576 |
| RAU      | 0.951155    | 0.951155       | 0.951155 | 0.393576 |
| GRAU     | 0.949846    | 0.949846       | 0.949846 | 0.393576 |
| PRES     | 0.685240    | 0.685240       | 0.685240 | 0.393576 |
| CAA      | 0.488512    | 0.488512       | 0.488512 | 0.393576 |
| PIR      | 1.010812    | 9.825e+04      | 9.825e+04 | <2e-16 |
| JAR      | 1.044996    | 2.225e+05      | 2.225e+05 | <2e-16 |

Residual standard error: 0.01609 on 938 degrees of freedom
Multiple R-squared: 1, Adjusted R-squared: 1

F-statistic: 6.326e+14 on 27 and 938 DF, p-value: < 2.2e-16

Estos son los resultados del error

IV. CONSEJOS UTILES.

PARA CONTINUAR CON LA DETERMINACIÓN DE CUÁL ES LA MEJOR ESTACIÓN QUE DEBE SER ATENDIDO CON LA DOTACIÓN DE TRENES ES LA ESTACIÓN GRAU CON UN VALOR DE 447, ESTE RESULTADO, DEBE SER COMPARADO CON LA MATRIZ.

REALIZAR ENTRENAMIENTOS NO PERMITE VISUALIZAR LOS SIGUIENTES DATOS

En los siguientes resultados se aprecia los resultados obtenidos de la ecuacion Ridge:
VI. DETERMINACION DE LAS MATRICES PARA TOMA DE DECISIONES ESTRUCTURADAS.

<table>
<thead>
<tr>
<th>Criterios de decisión</th>
<th>Cod</th>
<th>Estaciones del Tren</th>
<th>Requerir</th>
<th>Precio</th>
<th>Tiempo</th>
<th>Transmitir SIT del tren por Redes Sociales</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comprar más trenes</td>
<td>1</td>
<td>San Borja Sur</td>
<td>8</td>
<td>800000</td>
<td>40</td>
<td>10000</td>
</tr>
<tr>
<td>Mejorar sentido flujo de los paraderos</td>
<td>2</td>
<td>2</td>
<td>7000000</td>
<td>35</td>
<td>12000</td>
<td></td>
</tr>
<tr>
<td>El tiempo de espera</td>
<td>3</td>
<td>Grau</td>
<td>5</td>
<td>500000</td>
<td>30</td>
<td>6000</td>
</tr>
<tr>
<td>Realizar recargas virtuales</td>
<td>4</td>
<td>5</td>
<td>1000000</td>
<td>10</td>
<td>276.67</td>
<td></td>
</tr>
<tr>
<td>Transmitir SIT del tren por Redes Sociales</td>
<td>6</td>
<td>6</td>
<td>1000000</td>
<td>10</td>
<td>276.67</td>
<td></td>
</tr>
</tbody>
</table>

A. DETERMINACION DE COEF. PEARSON

Las Series temporales, generan el Coeficiente de Determinación Pearson ($R^2$) de 0.98 en Figura 01, sin embargo en la Figura 2 se aprecia que la linealidad no tiene la capacidad de comportarse en una variación no lineal, y es menos confiable por darnos un Coeficiente de Determinación de 0.5577.

V. ÁRBOL DE DECISIONES.

AQUÍ NOS ILUSTRA QUE LA ESTACION NARANJAL ES EL QUE MAS PREDOMINA EN EL TRAMO ANALIZADO

Ecuación Empleada:

En la ecuación 1, se aprecia el modelo matemático de predicción:

$ y = \beta_0 + \beta_1 VES + \beta_2 PIN + \beta_3 PUM + \beta_4 VMA + \beta_5 MAU + \beta_6 SJU + \beta_7 ATO + \beta_8 ICH + \beta_9 AYA + \beta_{10} CAB + \beta_{11} ANG + \beta_{12} SBS + \beta_{13} CULT + \beta_{14} NAR + \beta_{15} GAM + \beta_{16} BAY$

(1)

$f(x_i) = \beta_0 + \sum_{i=1}^{p} \beta_i x_i$

(II)

Donde:

$ y = \text{Variable a predecir}$

$x = \text{Tabla de Datos}$
VII. Conclusiones:

Se concluye de la figura 04, que no es necesaria emplear varios datos de ingreso, lo más importante es tener los de mayor incidencia en la toma de decisiones planeadas.

Los resultados obtenidos en las Series Temporales han permitido tomar mejores criterios a los pesos que se asignan para la Toma de decisiones Estructuradas y no Estructurada.

Las series temporales son más versátiles en no lineal a comparación de las ecuaciones pre determinadas en el programa Excel.

Se debe atender a la Estación Grau con la adquisición de más vagones, mejorando la ubicación de la dirección de los contra flujos que se presentan en paraderos luego se debe de atender a la estación San Borja Sur.

El modelo matemático que ayuda a la predicción es:

\[ y = \beta_0 + \beta_1 \text{VES} + \beta_2 \text{PIN} + \beta_3 \text{PUM} + \beta_4 \text{VMA} + \beta_5 \text{SJIU} + \beta_6 \text{ATO} \\
+ \beta_7 \text{ICH} + \beta_8 \text{AYA} + \beta_9 \text{CAB} + \beta_{10} \text{ANG} + \beta_{11} \text{SBS} + \beta_{12} \text{CULT} \\
+ \beta_{13} \text{NAR} + \beta_{14} \text{GAM} + \beta_{15} \text{BAY} \]

VIII. Otras Recomendaciones

Debe existir Buena correlación entre los datos a emplear

IX. BIBLIOGRAFÍA


