

Forecasting

Primera parte

José Burgos

2023-12-29

Los pronósticos se pueden aplicar a diversos tipos de series de tiempo, la robustez de nuestros pronóstico van a depender de:

1. Que tanto conocemos de los factores que contribuyen a ellos.
2. La cantidad de datos disponibles.
3. Cuán parecido es el futuro al pasado.
4. Si los pronósticos pueden afectar lo que estamos tratando de pronosticar.

pronóstico, objetivos y planificación

En la practica se suele confundir los conceptos de pronósticos, objetivos y la planificación, aunque estos estén relacionados, son diferentes entre si.

pronóstico: Enfocado en predecir el futuro lo más realista posible, utilizando datos históricos y conocimiento sobre eventos futuros que podrían afectar lo pronosticado.

Objetivos: Estos están asociados a lo que nos gustaría obtener en el futuro, aunque estos deberían estar anclados a los pronósticos para así ser mas realistas, pero suelen estar distanciados.

Planificación: Es la respuesta a pronósticos y objetivos. La planificación implica determinar las acciones apropiadas para que sus pronósticos coincidan con sus objetivos.

Determinar que pronosticar

Antes de iniciar un pronóstico se deben de conocer con certeza:

- Frecuencias de los datos.
- Tiempo con antelación que se requieren los pronósticos.
- Horizontes de tiempo que se deben pronosticar.

Es importante dedicar tiempo a hablar con las personas que utilizarán los pronósticos para asegurare de comprender sus necesidades y cómo se utilizarán los pronósticos.

Datos y métodos de pronósticos

Si no hay datos disponibles o estos no son relevantes, se deben utilizar métodos de pronósticos **cualitativos**

La **previsión cuantitativa** se puede aplicar cuando se cumple dos condiciones:

1. Se dispone de información numérica sobre el pasado.
2. Es razonable suponer que algunos aspectos de los patrones pasado continuarán en el futuro.

La mayoría de los problemas de predicción *cuantitativa* utilizan datos de **series de tiempo** (recopilados a intervalos regulares a lo largo del tiempo) o **datos transversales** (recopilados en un único momento).

Pronósticos de series de tiempo

Ejemplos de datos de series de tiempo:

- Beneficios anuales de Google.
- Resultados de ventas trimestrales de Amazon.
- Ventas minoristas semanales.
- Precios diarios de las acciones de **IBM**.
- Demanda de electricidad por hora.

Todo lo que se observa secuencialmente a lo largo del tiempo es una serie de tiempo. Algunos intervalos de tiempos o frecuencia más regulares: hora, día, semana, mes, trimestre, año.

Al pronosticar datos de series temporales, el objetivo es estimar cómo continuará la secuencia de observaciones en el futuro.

Los métodos más simples de pronóstico de series de tiempo utilizan sólo información sobre la variables que se va a pronosticar y no intentan descubrir los factores que afectan su comportamiento.

Los métodos de descomposición son útiles para estudiar la tendencia y los patrones estacionales en una serie temporal.

Variables predictoras y pronóstico de series temporales.

Las variables predictoras suelen ser útiles en el pronóstico de series temporales. Por ejemplo, supongamos que deseamos pronosticar la demanda de electricidad (ED) por hora de una región cálida durante el período de verano. Un modelo con variables predictoras podría tener la forma:

$$DE = F(\text{temperatura actual}, \text{fortaleza de la economía}, \text{población}, \text{hora del día}, \text{día de la semana}, \text{error})$$

La relación no es exacta: siempre habrá cambios en la demanda de electricidad que las variables predictoras no pueden explicar. El término de “error” de la derecha permite la variación aleatoria y los efectos de variables relevantes que no están incluidas en el modelo. A esto lo llamamos **modelo explicativo** porque ayuda a explicar qué causa la variación en la demanda de electricidad.

Debido a que los datos de demanda de electricidad forman una serie de tiempo, también podríamos usar un **modelo de serie de tiempo** para realizar pronósticos. En este caso, una ecuación de pronóstico de series de tiempo adecuada es de la forma:

$$DE_{t+1} = F(DE_t, DE_{t-1}, DE_{t-2}, DE_{t-3}, \dots, error),$$

dónde t es la hora actual, $t+1$ es la próxima hora, $t-1$ es la hora anterior, $t-2$ fue hace dos hora, y así sucesivamente. En este caso, la predicción del futuro se basa en valores pasados de una variable, pero no en variables externas que puedan afectar al sistema. También existe un tercer tipo de modelo que combina las características de los dos modelos anteriores. Por ejemplo:

$$DE_{t+1} = F(DE_t, temperatura\ actual, hora\ del\ día, día\ de\ la\ semana, error)$$

A este tipo de “modelos mixtos” se les han dado varios nombres en diferentes disciplinas. Se conocen como modelos de regresión dinámica, modelos de datos de panel, modelos longitudinales, modelos de función de transferencia y modelos de sistemas lineales (suponiendo que \mathbf{F} es lineal).

Un modelo explicativo es útil porque incorpora información sobre otras variables, en lugar de sólo valores históricos de la variable que se va a pronosticar. Sin embargo, hay varias razones por las que el pronosticador podría seleccionar un modelo de serie temporal en lugar de un modelo explicativo o mixto:

1. *El sistema puede no entenderse, e incluso si se entendiera, puede resultar extremadamente difícil medir las relaciones que se supone gobiernan su comportamiento.*
2. *Es necesario conocer o pronosticar los valores futuros de los diversos predictores para poder pronosticar la variable de interés, y esto puede resultar demasiado difícil.*
3. *La principal preocupación puede ser sólo predecir lo que sucederá, no saber por qué sucede.*
4. *El modelo de series temporales puede ofrecer pronósticos más precisos que un modelo explicativo o mixto.*

Los pasos básicos para realizar un pronóstico

1 Definición del problema.

Definir cuidadosamente el problema requiere comprender la forma en que se utilizarán los pronósticos, quién los requiere y cómo encaja la función de pronóstico dentro de la organización que requiere los pronósticos.

Recopilación de información.

Siempre se requieren al menos dos tipos de información: (a) datos estadísticos y (b) la experiencia acumulada de las personas que recopilan los datos y utilizan los pronósticos. A menudo será difícil obtener suficientes datos históricos para poder ajustar un buen modelo estadístico. En ese caso, se pueden utilizar los métodos de pronóstico de juicio. En ocasiones, los datos antiguos serán menos útiles debido a cambios estructurales en el sistema que se pronostica; entonces podemos optar por utilizar solo los datos más recientes. Sin embargo, recuerde que los buenos modelos estadísticos manejarán los cambios evolutivos en el sistema; No desperdicié información valiosa innecesariamente.

Análisis preliminar (exploratorio).

Comience siempre graficando los datos. ¿Existen patrones consistentes? ¿Existe una tendencia significativa? ¿Es importante la estacionalidad? ¿Existe evidencia de la presencia de ciclos económicos? ¿Hay valores atípicos en los datos que deban ser explicados por quienes tienen conocimientos expertos? ¿Qué tan fuertes son las relaciones entre las variables disponibles para el análisis?

Elección y ajuste de modelos.

El mejor modelo a utilizar depende de la disponibilidad de datos históricos, la solidez de las relaciones entre la variable de pronóstico y cualquier variable explicativa, y la forma en que se utilizarán los pronósticos. Es común comparar dos o tres modelos potenciales. Cada modelo es en sí mismo una construcción artificial que se basa en un conjunto de supuesto (explícitos e implícitos) y normalmente implica uno o más parámetros que deben estimarse utilizando los datos históricos conocidos.

Usar y evaluar un modelo de pronóstico.

El rendimiento del modelo sólo puede evaluarse adecuadamente después de que los datos para el período de pronóstico estén disponibles. Cuando se utiliza un modelo de pronóstico en la práctica, surgen numerosos problemas prácticos, como cómo manejar los valores faltantes y los valores atípicos, o cómo lidiar con series temporales cortas.

La perspectiva de la previsión estadística

Lo que intentamos pronosticar es desconocido (de lo contrario no lo estaríamos pronosticando), por lo que podemos considerarlo como una variable aleatoria.

En la mayoría de las situaciones de pronóstico, la variación asociada con lo que estamos pronosticando se reducirá a medida que se acerque el evento. En otras palabras, cuanto más adelante pronosticamos, más inciertos somos.

Cuando obtenemos un pronóstico, estamos estimando la mitad del rango de valores posibles que podría tomar la variable aleatoria. A menudo, un pronóstico va acompañado de un **intervalo de predicción** que proporciona un *rango* de valores que la variable aleatoria podría tomar con una probabilidad relativamente alta. Por ejemplo, un intervalo de predicción del 95% contiene un rango de valores que debe incluir el valor futuro real con una probabilidad del 95%.

Su utiliza el sub-índice t para el tiempo. Por ejemplo, y_t denotará la observación en el momento t . Supongamos que denotamos toda la información que hemos observado como Γ y queremos pronosticar y_t dado lo que sabemos en Γ . El conjunto de valores que podría tomar esta variable aleatoria, junto con sus probabilidades relativas se conoce como “distribución de probabilidad” de $y_t|\Gamma$. En pronóstico, a esto lo llamamos como **distribución de pronóstico**.

Cuando hablamos de **pronóstico**, generalmente nos referimos al valor promedio de la distribución del pronóstico, y le ponemos un **sombrero** a y para mostrar esto. Así, escribimos el pronóstico de y_t como \hat{y}_t , es decir, el promedio de los valores posibles que y_t podría tomar dado todo lo que sabemos.

A menudo resulta útil especificar exactamente qué información hemos utilizado para calcular el pronóstico. Luego escribiremos, por ejemplo, $\hat{y}_{t|t-1}$ significa el pronóstico de y_t teniendo en cuenta todas las observaciones anteriores (y_1, \dots, y_{t-1}). Similarmente, $\hat{y}_{t+h|t}$ significa el pronóstico de y_{t+h} teniendo en cuenta y_1, \dots, y_t (es decir, un h -Previsión de pasos teniendo en cuenta todas las observaciones hasta el momento t).

Ejercicios

1. Para los casos 3 y 4 de la sección 1.5, enumere las posibles variables predictoras que podrían ser útiles, suponiendo que los datos relevantes estén disponibles.

Caso 3 Una gran empresa de flotas de automóviles nos pidió que le ayudáramos a pronosticar los valores de reventa de los vehículos. Comparan vehículos nuevos, los alquilan por tres años y luego los venden. Mejores pronósticos de valores de venta de vehículos significarían un mejor control de las

ganancias; Comprender qué afecta los valores de reventa puede permitir que se desarrollen políticas de arrendamiento y ventas para maximizar las ganancias.

En aquel momento, un grupo de especialistas pronosticaba los valores de reventa. Desafortunadamente, vieron cualquier modelo estadístico como una amenaza para sus trabajos y no cooperaron a la hora de proporcionar información. Sin embargo, la empresa proporcionó una gran cantidad de datos sobre vehículos anteriores y sus eventuales valores de reventa.

Caso 4

En este proyecto, necesitábamos desarrollar un modelo para pronosticar el tráfico aéreo semanal de pasajeros en las principales rutas nacionales de una de las principales aerolíneas de Australia. La empresa requería previsiones del número de pasajeros para cada ruta nacional importante y para cada clase de pasajero (clase económica, clase ejecutiva y primera clase). La empresa proporcionó datos de tráfico semanales de los seis años anteriores.

El número de pasajeros aéreos se ve afectado por las vacaciones escolares, los principales eventos deportivos, las campañas publicitarias, el comportamiento competitivo, etc. Las vacaciones escolares a menudo no coinciden en diferentes ciudades australianas y los eventos deportivos a veces se trasladan de una ciudad a otra. Durante el periodo de los datos históricos se produjo una importante huelga de pilotos durante la cual no hubo tráfico durante varios meses. También se lanzó y cerró una nueva aerolínea de precio reducido. Hacia el final de los datos históricos, la aerolínea había probado una redistribución de algunos asientos de clase económica a clase ejecutiva y algunos asientos de clase ejecutiva a primera clase. Sin embargo, después de varios meses, las clasificaciones de asientos volvieron a la distribución original.

Respuestas:

Caso 3

Variables predictoras para obtener los valores de reventa de los vehículos rentados:

- Kilometraje del vehículo.
- Año de fabricación del vehículo.
- Precio promedio del vehículo en el mercado.
- Precio de adquisición del vehículo.
- Tipo de combustible del vehículo (gasoil, gasolina, gas, eléctrico).

$$\text{valor_reventa} = \beta_0 + \beta_1 \text{ km} + \beta_2 \text{ year_fabr} + \beta_3 \text{ mean_veh} + \beta_4 \text{ price_vehiculo} + \beta_5 \text{ combustible} + \epsilon$$

2. Para el caso 3 en la sección 1.5, describa los cinco pasos del pronóstico en el contexto de este proyecto.