

Prediciendo indicadores en el retail





NELSON BALOIAN

Profesor Asociado del Departamento de Ciencias de la Computación de la Universidad de Chile, y profesor visitante regular en las universidades de Waseda, Japón, y de Duisburg-Essen, Alemania. PhD en Ciencias por la Universidad de Duisburg, Alemania. Sus áreas de interés de investigación han sido los sistemas de apoyo computacionales para el aprendizaje, sistemas distribuidos y machine learning. Ha sido autor de más de 30 artículos en revistas indexadas y más de 100 en conferencias internacionales.

nbaloian@dcc.uchile.cl



JOSÉ A. PINO

Profesor Titular del Departamento de Ciencias de la Computación de la Universidad de Chile. Cofundador del DCC, ha servido como Presidente de la Sociedad Chilena de Ciencias de la Computación y Presidente de CLEI. Sus áreas de interés actuales son aprendizaje de máquina y administración de procesos de negocio. Su investigación se ha publicado en *journals*, incluyendo Expert Systems with Applications, Information Systems Frontiers y ACM Computing Surveys.

jpino@dcc.uchile.cl



BELISARIO PANAY

Ingeniero Civil en Computación y Magíster en Ciencias mención Computación de la Universidad de Chile. Líneas de trabajo: aprendizaje automático. En Twitter lo encuentras como @belisariops.

bpanay@dcc.uchile.cl

Las tiendas no-online en nuestro país han sufrido fuerte con la pandemia del COVID-19. Como han debido permanecer cerradas durante largos periodos, en éstos no han recibido ingresos, pero sí han debido afrontar sus gastos fijos. En consecuencia, es importante para ellas optimizar su operación cuando las condiciones mejoren. ¿Cómo podría un administrador conocer su situación en cuanto a ventas?

Un primer indicador de la eficacia de las ventas es averiguar la cantidad de personas que visita la tienda por unidad de tiempo. En caso de que nadie entre a la tienda, mal puede haber ventas. Por el contrario, si muchas personas van a la tienda, hay mayor probabilidad de que compren. Entonces, el número de personas que entran a la tienda por unidad de tiempo (*foot traffic*) es un indicador

de la eficacia. Pero podría darse que muchas personas visiten la tienda, pero pocas compren, así es que la proporción de personas que compran con respecto al total de visitantes (*conversion rate*) es un segundo indicador apropiado. Un tercer indicador es el total de ventas en dinero realizadas por unidad de tiempo.

Estos indicadores no sólo pueden servir para conocer la situación pasada, sino que pueden usarse para predecir el desempeño futuro. En la medida que eso sea posible, el administrador puede prepararse para esa futura demanda. Así, puede preparar el número suficiente de vendedores y cajeros, el stock de productos a vender y el flujo de caja esperado.

En este artículo presentamos un proyecto de desarrollo de software de pre-

dicción de indicadores para tiendas del retail llevado a cabo por nuestro grupo MARAL (Machine learning Research Applied Lab) con financiamiento de Corfo Innova y apoyo de la empresa Follow Up. El grupo está compuesto, además de los autores del artículo, por Sergio Peñafiel (estudiante de Magíster, ya graduado), Jonathan Frez (candidato a Doctorado) y Cristóbal Fuenzalida (estudiante de Ingeniería Civil en Computación).

El problema

Un enfoque al tema de la predicción de indicadores es usar información previa sobre estos mismos indicadores, pero ¿cómo obtenerla? Aquí viene el aporte de la empresa Follow Up. Esta compañía



nacional ha instalado cámaras a la entrada de gran número de tiendas de Chile, Colombia, Perú, Japón y otros países. Con ayuda de ellas, ha construido una base de datos del *foot traffic* de esas tiendas: la cantidad de visitantes por hora. Los datos de los otros indicadores se pueden obtener del software de ventas de las tiendas mismas.

La predicción que se pretende debe cumplir dos requerimientos. Por una parte, el software debe proveer la importancia que cada variable de entrada tiene en la resultante salida. Esto porque sabiendo ese peso, los administradores pueden reaccionar mejor a cambios inesperados de las condiciones ambientales, tales como un nuevo día festivo, o una campaña de marketing. El segundo requerimiento se refiere al grado de confianza o por el contrario, de incertidumbre, de la predicción. Estos requerimientos implican que el modelo a usarse debe tener algún grado de transparencia, en oposición a los modelos de “caja negra”, que no entregan ninguna información adicional a la predicción misma. Normalmente, los modelos de caja negra entregan predicciones con menos error que los métodos transparentes, por lo que típicamente se paga un precio en términos de calidad de predicción si se quiere transparencia. En nuestra investigación, entonces, intentamos desarrollar un método de tipo transparente que tuviera similar calidad de predicción que los modelos de caja negra.

Metodología

El método usado para encontrar un modelo satisfactorio fue el siguiente. Después de investigar los modelos existentes, se trató de mejorar alguno promisorio. En seguida, se trabajó en la *embedding*, es decir, investigar la mejor manera de codificar la información disponible (en un vector) para que sea entendida por el modelo. Pos-

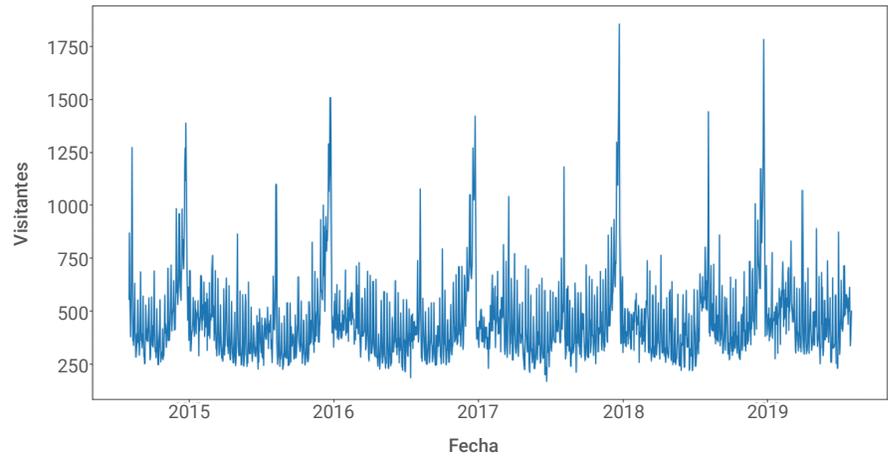


Figura 1. Número de visitantes a una tienda por hora de acuerdo a la fecha.

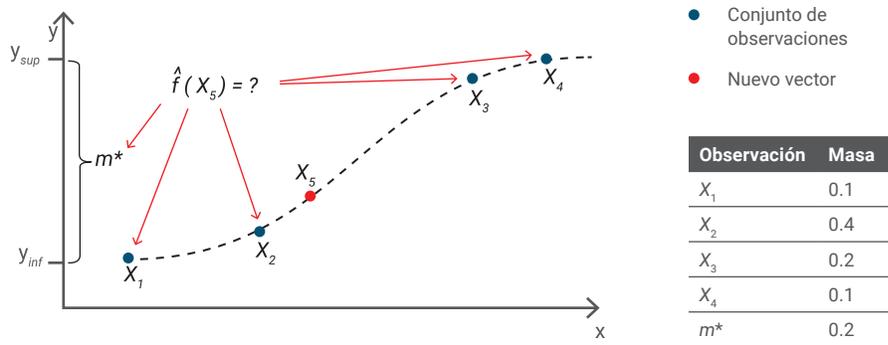


Figura 2. Ilustración del método.

teriormente, se realizaron experimentos con el modelo estudiando cómo se desempeña con nuestros datos. En cada iteración, se cambiaba el *embedding* o el modelo, hasta obtener resultados satisfactorios.

Los datos utilizados en nuestro estudio no fueron todos los disponibles. Se buscaron datos de tiendas con datos completos durante cuatro años. Sólo 20 tiendas cumplían estos requisitos, con información entre agosto de 2015 y el mismo mes de 2019. La Figura 1 muestra un típico patrón de comportamiento del *foot traffic* en dicho período.

El modelo desarrollado

Después de revisar varios modelos de predicción publicados en la literatura, seleccionamos para mayor investigación el modelo de *Evidential Regression* (EVREG), basado en una extensión difusa de las funciones de creencia, desarrollado por Petit-Renaud y Denoeux, publicado en 2004 [1]. Las funciones de creencia, a su vez, son parte de la Teoría de Dempster-Shafer también conocida como la teoría de evidencia [2]. Este método predice un valor usando un



[Nuestro enfoque] logra un buen desempeño en general, obteniendo resultados comparables a los mejores métodos probados en la literatura.

conjunto de observaciones pasadas. Al predecir, a cada observación se le asigna una masa que representa la similitud con el vector que se va a predecir. Luego estas masas son transformadas en una distribución de probabilidades con la que se calcula un valor esperado. En términos simples podemos explicarlos con la Figura 2.

Acá tenemos un ejemplo donde se tiene un conjunto de observaciones con 4 vectores (en azul). Éste sería el conjunto de entrenamiento de un problema. Y lo que se necesita es predecir la salida de una nueva observación (en rojo). Para encontrar el valor estimado de salida de esta nueva observación, se puede suponer que un vector que esté cerca de ella va a tener una salida parecida. En este ejemplo el segundo vector es el más cercano, así que se puede suponer que éste tiene la salida más parecida. Para reflejar esto, se le asigna una “masa” o importancia a cada uno de los puntos en el conjunto de observaciones según una función de distancia (por ejemplo, la euclidiana). Mientras mayor es la masa más importante es el punto. Esto significa que mientras más cerca está un punto de la nueva observación, más importante es la evidencia que éste entrega. Lo anterior, es similar a un regresor de k -vecinos más cercanos donde se calcula una contribución según una distancia. Pero la teoría de la evidencia fue creada para razonar con incerteza. Por ejemplo, puede ocurrir que los puntos del conjunto de entrenamiento están muy lejos de la nueva observación, por lo que no se puede estar seguro de la respuesta que se entrega. Para esto se agrega otra fuente de información, en este ejemplo además de las observaciones de entrenamiento se conocen cuáles son los valores por los que se mueve la salida de estas observaciones. Para

reflejar la incertidumbre de la respuesta, *Evidential Regression* entrega una masa a esta observación del intervalo de salida, la cual representa el grado de incertidumbre del proceso.

Evidential Regression calcula las masas de cada una de las observaciones en el set de entrenamiento con una distancia usando un vector de características. Este vector de características codifica la información de un problema, por ejemplo, para este problema las dimensiones del vector tenían datos como el trimestre, mes, día del mes y día de la semana del evento que se quería predecir. Al calcular la distancia entre estos vectores de características *Evidential Regression* supone que todas las dimensiones o características de los vectores de entrenamiento son igual de importantes para calcular la similitud o importancia entre vectores. ¿Pero qué pasa si estamos prediciendo *foot traffic* para una tienda que se encuentra cerca de oficinas? Se esperaría que los días de semana afecten mucho la predicción de sus entradas, ya que es en días hábiles cuando más público se observa. Entonces una diferencia de una dimensión como el día de la semana no afectará de la misma manera que una diferencia en una dimensión como el trimestre en que se encuentra el evento. Entonces lo que proponemos es una versión mejorada de *Evidential Regression* que llamaremos *Weighted Evidential Regression* (WEVREG) que usa una distancia ponderada, para el cálculo de masas [3]. Estos pesos serán aprendidos durante la fase de entrenamiento del algoritmo usando descenso de gradiente. Cada peso representa la importancia de la dimensión del vector de entrada para predecir las salidas del modelo y así ayudará al modelo a aumentar su poder de predicción. Además, para disminuir la complejidad del algoritmo,

Método	RMSE
RF	0.1041 ± 0.01
WEVREG	0.1088 ± 0.01
SVM	0.1133 ± 0.01
GP	0.1321 ± 0.01
LSTM	0.1422 ± 0.02
SARIMA	0.1489 ± 0.03

Figura 3. Resultados.

se usará una k -NN (*k-nearest neighbors*) donde sólo se calcularán las masas de los k vecinos más cercanos.

Evaluación

Los datos consisten en datos de tres indicadores *foot traffic*, *conversion rate* y total de ventas de 20 tiendas, entre agosto de 2015 y el mismo mes de 2019. La meta de este problema es predecir cada uno de estos indicadores para el mes de julio de 2019, usando todos los datos que se tienen disponibles. Para esto se tenían datos diarios de cada uno de estos indicadores. El vector de características se creó desagregando el tiempo de los eventos y se les aplicó una representación circular, por ejemplo, el día de la semana (representado con un entero que va de 0 a 6) se separa en el seno y coseno del día de la semana. A esto se agregó una secuencia que tenía los valores de salida de los últimos 6 meses con un intervalo de 1 mes, esto significa que para el día 1 de agosto, se tenían los valores de salida del 1 de julio, 1 de junio y así sucesivamente.

Para evaluar cuantitativamente el método propuesto, se le comparó con otros métodos que han sido usados con anterioridad en la literatura. Estos son métodos como *Random Forest* (RF), *Long Term Short Memory* (LSTM), *Support*

Otra característica destacable [de nuestro enfoque] es su interpretabilidad [...]: el modelo da los pesos de los atributos después del entrenamiento.

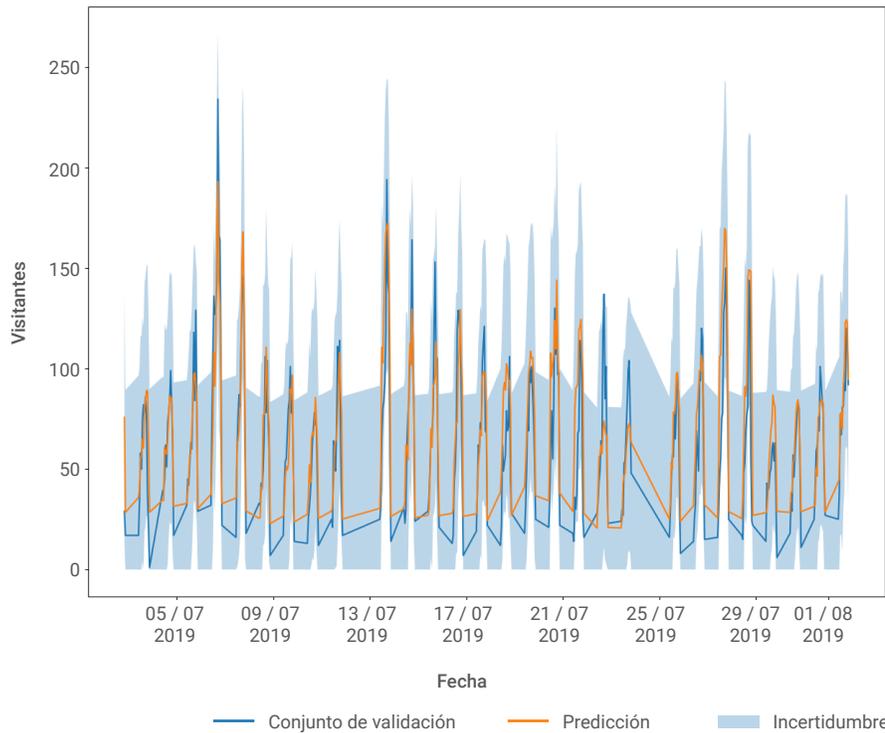


Figura 4. Predicción del *foot traffic*.

Vector Machine (SVM), Gaussian Process (GP) y Seasonal Autoregressive Integrated Moving Average (SARIMA). En la Figura 3 pueden observarse los resultados. El error medido fue el *Root Mean Squared Error* (RMSE), mientras menor es este error mejor es el método. De todos los métodos puestos a prueba el RF fue el que obtuvo los mejores resultados, pero fue seguido de cerca por nuestro método propuesto, el cual obtuvo el segundo lugar para nuestro conjunto de datos.

En la Figura 4 podemos observar la predicción del *foot traffic* para una tienda en particular. En color azul se ve la curva real (determinada a partir de las observaciones de las cámaras) y en naranja se ve la

predicción de nuestro modelo. Como se puede ver, se ajusta bastante a los datos reales y además entrega un intervalo de variación el cual es la incertidumbre de la respuesta. Las predicciones son calculadas como un valor esperado, cuando se predice una nueva observación, por ejemplo la del día 5 de julio, se calculan las masas (o importancias) de cada uno de sus *k* vecinos. Para llegar a una predicción se multiplican estas masas por los valores de salida de sus vecinos y, además, se agrega el término de incertidumbre el cual se calcula como la masa de la incertidumbre por el valor medio del intervalo en que se mueven los vecinos de la nueva observación. Esto se muestra en la Ecuación (1), donde *N* es la cantidad total de vecinos,

x es la nueva observación, *m_i* es la masa del vecino *i*, *y_i* es el valor de salida del vecino *i*, *m** es la masa de la incertidumbre, *sup y* y *inf y* son el valor máximo y mínimo de salida del conjunto de vecinos.

$$\hat{y} = \sum_{i=1}^N m_i(x) \cdot y_i + \frac{m^*(x) \cdot (\sup_{y \in \mathcal{L}} y + \inf_{y \in \mathcal{L}} y)}{2} \tag{1}$$

Con esto se puede calcular un límite superior e inferior de esta predicción como se muestra en las Ecuaciones 2 y 3.

$$\hat{y}^* = \sum_{i=1}^N m_i(x) \cdot y_i + m^*(x) \cdot \sup_{y \in \mathcal{L}} y \tag{2}$$

$$\hat{y}_* = \sum_{i=1}^N m_i(x) \cdot y_i + m^*(x) \cdot \inf_{y \in \mathcal{L}} y \tag{3}$$

Además de entregar una predicción y su intervalo de incertidumbre, el método es capaz de entregar una medida que representa la importancia de cada una de las variables de la entrada que es usada para predecir. En la Figura 5 se puede observar la importancia de cada una de estas variables. Como se puede ver, para esta tienda en particular el valor más importante para la predicción es la secuencia de valores anteriores, en especial, los valores del *foot traffic* registrados hace 1 mes y 6 meses antes del día que se quiere predecir.

Conclusiones

Del análisis de los resultados presentados en la sección anterior podemos ver que nuestro enfoque logra predecir correctamente los indicadores claves del retail. En efecto, en la Figura 4, podemos ver que en general las predicciones están bien

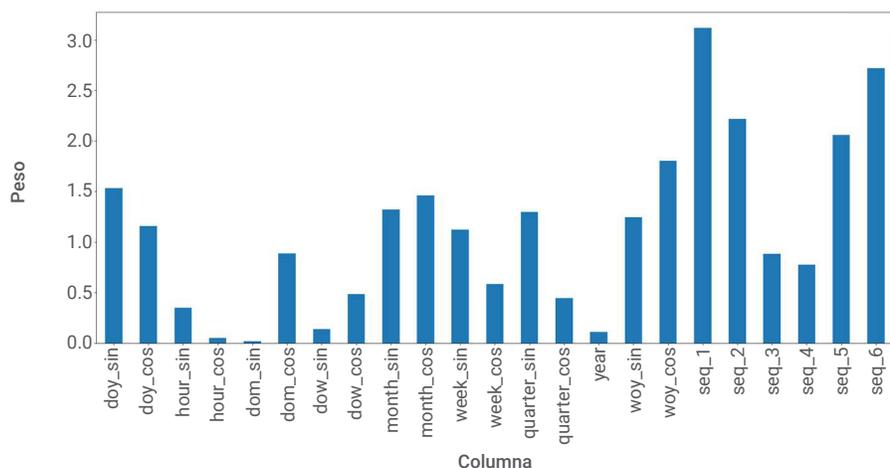


Figura 5. Importancia de los parámetros.

ajustadas a las curvas reales. El método no tiene problemas para detectar las puntas y valles de los valores reales, aunque no alcanza los mismos valores superiores e inferiores. En particular, el modelo apenas alcanza los valores extremos de la predicción. Esta deficiencia puede explicarse por la naturaleza de la predicción con k -NN; hay que tener en cuenta que para predecir un valor 0 (el mínimo en nuestro caso), el modelo requiere que todos los vecinos que observa también deben tener el valor 0; si alguno de ellos no tiene un valor 0 entonces “mueve” la predicción hacia el centro.

Otra característica de WEVREG es su capacidad de proporcionar intervalos de

variación. En la misma figura, podemos observar que casi todos los valores reales están dentro del intervalo de variación. Sin embargo, podemos observar que el ancho de este intervalo es amplio, cubriendo alrededor del 30% del rango de predicción. Esto podría deberse a que los vectores utilizados por los métodos posteriores al enfoque k -NN para cada predicción no son muy similares entre sí, obteniendo una alta incertidumbre para el proceso.

Otra característica destacable del modelo WEVREG es su interpretabilidad. Como se muestra en la Figura 5, el modelo da los pesos de los atributos después del entrenamiento. En este caso, el mode-

lo está utilizando un *embedding* cíclico junto con la secuencia de los seis meses anteriores para una tienda en particular. A partir de esta figura, queda claro que para esta tienda en particular algunos componentes como el año o componentes parciales de la hora y el día del mes no son realmente importantes para predecir sus visitantes. Además, las características más importantes para predecir el número de visitantes parecen ser el número de visitantes observados en meses anteriores. El mes anterior es el más importante, y posteriormente se observa una disminución de importancia seguida de un aumento en el quinto y sexto mes que podría deberse al comportamiento cíclico de los visitantes de esta tienda en particular.

Hablando de desempeño, WEVREG logra un buen desempeño en general, obteniendo resultados comparables a los mejores métodos probados en la literatura [4]. Como se muestra en las Figura 3, RF obtiene el mejor desempeño y este resultado coincide con lo reportado previamente en la literatura [5].

Como se trata de problemas de series de tiempo, se esperaba que LSTM, que es un método basado en aprendizaje profundo, obtuviera los mejores resultados, pero no pudo superar nuestro método propuesto en general. Una posible razón detrás de este bajo rendimiento podría ser el uso de una arquitectura de red única para todas las tiendas. ■

REFERENCIAS

- [1] S. Petit-Renaud, T. Denœux: Nonparametric regression analysis of uncertain and imprecise data using belief functions. *Int. J. Approx. Reason.* 35, 2004, 1–28.
- [2] G. Shafer: Dempster’s rule of combination. *Int. J. Approx. Reason.* 79, 2016, 26–40.
- [3] B. Panay, N. Baloian, J.A. Pino, S. Peñafiel, H. Sanson, N. Bersano: Feature selection for health care costs prediction using Weighted Evidential Regression. *Sensors* 20(16), 2020, 4392.
- [4] B. Panay, N. Baloian, J.A. Pino, S. Peñafiel, J. Frez, C. Fuenzalida, H. Sanson: Forecasting key retail performance indicators using interpretable regression. *Sensors* 21(5), 2021, 1874.
- [5] S. Abrishami, P. Kumar, W. Nienaber: Smart stores: A scalable foot traffic collection and prediction system. In *Industrial Conference on Data Mining*; Springer: Cham, Switzerland, 2017, 107–121.