

Previsión de ventas basado en un modelo híbrido de pronóstico gris y un modelo de red neuronal artificial

Kennya Jeamileth Trochez Aguilar, Rodolfo Eleazar Pérez Loaiza,
Perfecto Malaquías Quintero Flores

Tecnológico Nacional de México,
Instituto Tecnológico de Apizaco,
Departamento de Estudios de Posgrado e Investigación,
México

{kwow1820, reploaiza, parafait}@gmail.com

Resumen. Los pronósticos de ventas son muy importantes para el sector minorista, ya que los pronósticos de ventas precisos ayudan a aumentar en gran medida los beneficios finales de estas. Las técnicas de planificación logística basadas en algoritmos heurísticos han sido favorables en comparativa a los métodos tradicionales de pronóstico, especialmente en los casos donde no hay datos históricos o bien, estos históricos son difíciles de interpretar. En esta propuesta se investiga la volatilidad de la demanda y otros factores desde fluctuaciones hasta cuellos de botella problemáticos para los ejecutivos de pymes. Debido a lo difícil de obtener un pronóstico preciso, se propone un método de pronóstico híbrido que fusiona una red neuronal artificial (RNA) y el modelo de predicción gris (GM (1,1)), donde el algoritmo obtiene las variables influyentes a través del análisis de correlación. Después de emplear la predicción gris y la RNA para predecir, se introduce la idea del MSE y se comparan los datos de ventas reales con el pronóstico gris junto con el resultado de la predicción de la red. Finalmente, la viabilidad y precisión de la predicción del algoritmo se verifican a través de los datos reales de las ventas. Los resultados experimentales muestran que, en la predicción de datos de ventas, el algoritmo tiene un error bajo comparado con otros métodos estadísticos y su rendimiento de predicción es mejor que el modelo ARIMA, que el modelo manual y que el modelo de series temporales.

Palabras clave: Red neuronal artificial, modelo gris, previsión de ventas, análisis de correlación, ARIMA, series temporales.

Sales Forecast Based on a Hybrid Gray Forecast Model and an Artificial Neural Network Model

Abstract. Sales forecasts are very important to the retail industry; as accurate sales forecasts help to greatly increase the bottom line of sales. Logistics planning techniques based on heuristic algorithms have been favorable compared to traditional forecasting methods, especially in cases where there are no historical data or where historical data are difficult to interpret. This proposal investigates demand volatility and other factors ranging from fluctuations to problematic

bottlenecks for SME executives. Due to the difficulty of obtaining an accurate forecast, a hybrid forecasting method is proposed that merges an artificial neural network (ANN) and the gray prediction model (GM (1,1)), where the algorithm obtains the influencing variables through correlation analysis. After using the gray prediction and ANN to predict, the idea of the MSE is introduced and the actual sales data is compared to the gray forecast along with the network prediction result. Finally, the feasibility and accuracy of the algorithm's prediction are verified through actual sales data. The experimental results show that, in the prediction of sales data, the algorithm has a low error compared to other statistical methods and its prediction performance is better than the ARIMA model, the manual model and the time series model.

Keywords: Artificial neural network, gray model, sales forecast, correlation analysis, ARIMA, time series.

1. Introducción

Hoy en día la ciencia de datos, junto con su análisis y las técnicas de aprendizaje automático y aprendizaje profundo están creciendo a nivel potencial y las empresas ahora buscan técnicas capaces de examinar, comprender e interpretar la información útil y ayudarlos a impulsar y mejorar la toma de decisiones para brindar un mejor servicio y una interpretación de la información más eficiente, mientras que el aprendizaje automático tiene sus orígenes en la informática como una de las ramas de la inteligencia artificial y uno de los componentes tecnológicos más importantes de la ciencia de datos.

Por otra parte, los pronósticos precisos son de gran importancia para el sector comercial, estos no sólo pueden proporcionar una base científica para la toma de decisiones comerciales y el trabajo de planificación de la empresa, sino que afectan directamente el rendimiento del mercado de productos de la empresa y las ganancias de ventas finales, además de que también ayudan a los comerciantes a ajustar de manera flexible sus sistemas de planificación y a mejorar la velocidad de respuesta a los cambios en la demanda, con el fin de obtener una ventaja competitiva y mayores beneficios económicos.

Los pronósticos erróneos dan lugar a productos fuera de venta o stocks innecesarios, lo que afectará directamente el crecimiento del comercio, e incluso provocará pérdidas significativas para el mismo. En el sector comercial minorista, hay muchos factores que afectan el desempeño de las ventas, como el comportamiento del cliente, los cambios estacionales y festivos, la inversión en publicidad de productos, las actividades de promoción comercial a gran escala e incluso algunas políticas económicas y comerciales. Estos factores de influencia provocan incertidumbre en la demanda de productos y presentan inconsistencias en la precisión de los pronósticos de ventas.

Actualmente, los métodos de pronóstico se dividen principalmente en dos tipos: uno se basa en la teoría matemática y estadística de métodos clásicos, como el suavizado exponencial, ARIMA, modelo de regresión o modelo de suavizado exponencial triple, el otro es un algoritmo heurístico moderno que utiliza una rama de la inteligencia artificial, como las redes neuronales artificiales (RNA). En realidad, la relación entre las ventas y los factores que influyen es bastante complicada, y los datos de ventas suelen ser asimétricos y cambian mucho.

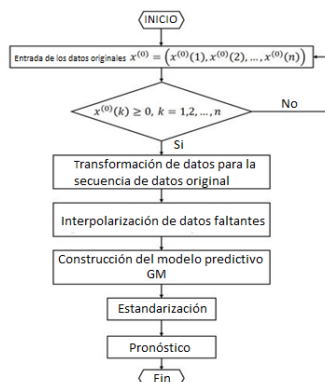


Fig. 1. Diagrama de flujo del funcionamiento del algoritmo de predicción gris (GM).

En los métodos clásicos, el suavizado exponencial, el modelo de regresión y el algoritmo ARIMA se basan en los modelos de Box y Jenkins, y estos dos métodos utilizan métodos lineales para el modelado de series de tiempo [2], pero el modelo lineal no puede capturar parte de la realidad, por lo que los problemas de ciclo asimétrico en la serie de tiempo son susceptibles a la irregularidad y variabilidad de los datos, por lo tanto es difícil obtener resultados de pronóstico precisos y no siempre es el método más apto para los pronósticos de ventas. Por el contrario, los algoritmos de redes neuronales generalmente pueden resolver eficazmente estos problemas y obtener resultados de predicción relativamente buenos [1,3].

Pero vale la pena señalar que la precisión de las predicciones de estos algoritmos depende en gran medida de sus objetivos experimentales, áreas de aplicación y si tienen una gran cantidad de datos históricos reales. En el caso de una cantidad limitada de datos históricos, es difícil obtener resultados de predicción ideales. En resumen, los algoritmos y los modelos de redes neuronales no siempre son ideales para obtener resultados satisfactorios al predecir ventas con datos limitados o bien con un excedente de datos anómalos, es por ello que en este trabajo se presenta una propuesta de una red neuronal que se adapta a casos donde el volumen de información útil es limitado.

2. Trabajos relacionados

En sus inicios, el empleo de las redes neuronales artificiales (RNA) para realizar pronósticos fue llevado a cabo por primera vez en el año 1964, cuando Hu empleó una RNA lineal adaptable de Widrow para el pronóstico del clima (Zhang, Patuwo & Hu, 1998). Pero no fue sino hasta 1986 cuando Rumelhart, Hinton & Williams (1986) introdujeron el algoritmo de retropropagación, que las RNA empezaron su verdadero desarrollo. Dos años más tarde Werbos (1988) reportó que la RNA entrenada por retropropagación “superó el desempeño de los métodos estadísticos tradicionales tales como los procedimientos de regresión y Box-jenkins en varios casos”.

En los primeros años del siglo XXI, las RNAs se popularizaron en el pronóstico de series de tiempo en diversas áreas que incluyen finanzas, generación de energía, medicina, recursos del agua y ciencias ambientales, entre otras (Maier & Dandy, 2000).

Otros estudios realizados por Smith & Gupta (2000) y Zhang (2004), mostraron aplicaciones de las RNA en problemas de investigación de operaciones, negocios y pronósticos de mercado.

En cuanto a investigaciones de estadísticas, existen diversas investigaciones sobre algoritmos de pronóstico de ventas, Sun [20] introdujo la aplicación de Extreme Learning Machine (ELM) en el pronóstico de ventas, y luego propusieron el algoritmo ELM extendido (EELM). Muchos experimentos han demostrado que el algoritmo es más estable. Hsu [8] propone el Modelo Gris (GM), que puede obtener mejores resultados de predicción en el caso de datos limitados.

3. Método teórico

Para resolver el problema presentado, esta investigación propone un algoritmo híbrido, que utiliza una combinación de algoritmo de modelo gris GM (1,1) y algoritmo de red neuronal y considera de manera integral la mayoría de los factores que pudieran influir en el resultado de las ventas totales reales.

En el proceso de previsión de las ventas, el algoritmo híbrido de este artículo resuelve el problema de la precisión de predicción inestable de los modelos alternos, esto gracias a la implementación de una red neuronal artificial. En el caso de un gran volumen de datos atípicos, o como cuando la cantidad de datos es limitada, los resultados de predicción son más fiables a diferencia de otros métodos. La prueba final muestra que la precisión de predicción de este algoritmo inteligente híbrido es mejor que la de ARIMA, series de tiempo y GM generalizado.

3.1. Modelo gris GM (1,1)

El modelo de pronóstico gris es un método de pronóstico que establece un modelo matemático y hace predicciones a través de una pequeña cantidad de información incompleta. La teoría del sistema gris es una teoría que estudia y resuelve el análisis, modelado, pronóstico, toma de decisiones y control del sistema gris.

La predicción gris es la predicción realizada en el sistema gris. En la actualidad, algunos métodos de predicción de uso común (como el análisis de regresión, etc.) requieren una muestra grande. Si la muestra es pequeña, a menudo causa grandes errores e invalida el objetivo de predicción. El modelo de pronóstico gris requiere menos información de modelado, operación conveniente y alta precisión de modelado. Es una herramienta eficaz para tratar con problemas de pronóstico de muestras pequeñas.

Suponga que la secuencia de datos original, $x^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\}$, al definir la ecuación 1:

$$x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i), k = 1, 2, 3, \dots, n, \quad (1)$$

Obtenemos una nueva serie $x^{(1)}(k)$: $x^{(1)}(k) = x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)$. Para algunos procesos, $x^{(1)}(k)$ es la solución de la ecuación 2 diferencial ordinaria formada en blanco:

$$\frac{dx^{(1)}}{dt} + \alpha x^{(1)} = u, \quad (2)$$

donde α y T son el número gris, es decir, los parámetros colgantes, que se calcula por el método de mínimos cuadrados. La ecuación (1) se llama GM (1, 1). La solución de (1) es:

$$x^{(i)}(k+1) = \left[x^{(0)}(1) - \frac{u}{a} \right] e^{-ak} + \frac{u}{a}, \quad (3)$$

La ecuación (2) se llama función de respuesta en el tiempo. Para $k \geq 2$, se llama fórmula de predicción, denotada por la ecuación 4:

$$x^{(\hat{0})}(k) = x^{(i)}(k+1) - x^{(i)}(k). \quad (4)$$

De lo anterior se puede ver que la esencia del modelo de predicción gris es generar mediante la acumulación de la secuencia de datos original una nueva generación, debilitar la influencia de factores de perturbación aleatorios, encontrar la ley de desarrollo de las cosas, generar una secuencia de datos con fuerte regularidad para posteriormente predecir el desarrollo futuro de la secuencia, porque el proceso se simula usando la curva exponencial.

Se debe mencionar que debido a que el modelo gris se genera acumulando secuencias de datos para fortalecer las leyes internas de los datos, en el caso de datos limitados, los resultados de la predicción son más precisos, ya que el método no requiere un estudio de grandes volúmenes de información. Además, es un modelo estable y adaptable ante la carencia de datos estadísticos, por lo que el algoritmo híbrido que se propone en este trabajo puede obtener resultados más precisos de predicción que otros algoritmos cuando la cantidad de datos es limitada.

3.2. Red neuronal artificial

Las redes neuronales artificiales (RNA), en el campo del aprendizaje automático, son una imitación de las redes neuronales biológicas, especialmente del modelo matemático o modelo de cálculo de la estructura y función del cerebro, utilizado para estimar o aproximar la función. En la mayoría de los casos, la red neuronal artificial puede cambiar su estructura interna sobre la base de información externa, es un sistema adaptativo y, en términos sencillos, tiene una función de aprendizaje. La red neuronal moderna es una herramienta de modelado de datos estadísticos no lineales.

Las RNAs generalmente se optimizan mediante un método de aprendizaje basado en matemáticas estadísticas, por lo que también es una práctica aplicación de la estadística. Por otro lado, en el campo de la inteligencia artificial, podemos tomar decisiones sobre la obtención de información mediante la aplicación de estadísticas matemáticas, es decir, a través de métodos estadísticos, las redes neuronales artificiales pueden tener capacidad de toma de decisiones y capacidad de juicio lógico similar a los humanos. Este método brinda una ventaja sobre el procesamiento de información y toma de decisiones.

3.3. Red neuronal Backpropagation (BPP)

Las redes BP tienen un método de entrenamiento supervisado. A la red se le presenta parejas de patrones, un patrón de entrada emparejado con un patrón de salida deseada. Por cada ejecución los pesos son ajustados de forma que disminuya el error entre la

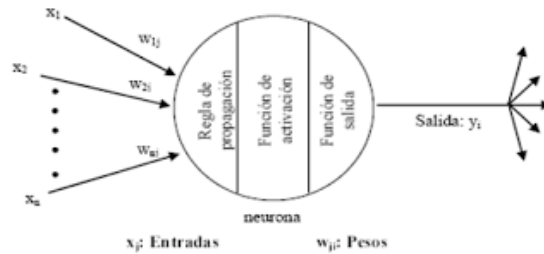


Fig. 2. Modelo del funcionamiento de una red neuronal artificial.

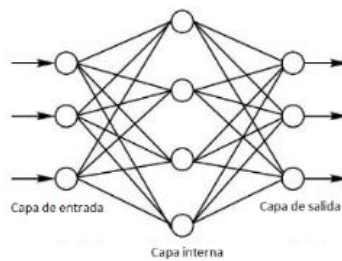


Fig. 3. Modelo arquitectónico de una red neuronal backpropagation (BP).

salida deseada y la respuesta de la red. Este algoritmo de aprendizaje conlleva una fase de propagación hacia adelante y otra fase de propagación hacia atrás. Ambas fases se realizan por cada patrón presentado en la sesión de entrenamiento.

3.3.1 Propagación hacia adelante

Esta fase de propagación hacia adelante se inicia cuando se presenta un patrón en la capa de entrada de la red. Cada unidad de la entrada se identifica con un elemento del vector patrón de entrada. Las unidades de entrada toman el valor de su correspondiente elemento del patrón de entrada y se calcula el valor de activación o nivel de salida de la primera capa.

Posteriormente, las demás capas realizarán la fase de propagación hacia adelante, el cual determina el nivel de activación de las otras capas. La unidad procesadora j obtiene la cantidad S_j según la siguiente ecuación:

$$S_j = \sum_i a_i w_{ji}, \tag{5}$$

Y genera la salida o nivel de activación empleando la siguiente ecuación:

$$salida = f(S_j). \tag{6}$$

La función f es una función umbral genérica, entre las que cabe destacar la función Sigmoide o la función Hiperbólica. El valor de la salida de la unidad j es enviado o transmitido a lo largo de todas las conexiones de salida de dicha unidad.

Cabe mencionar que las unidades procesadoras de la capa de entrada no realizan ninguna operación de cálculo con sus entradas, ni operaciones con funciones umbrales, sólo asumen su salida como el valor del correspondiente elemento del vector de entrada.

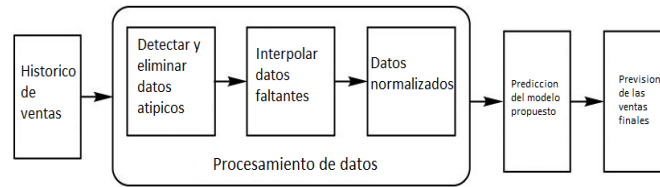


Fig. 4. Esquema del modelo híbrido propuesto.

Por otro lado, algunas redes BP utilizan unidades llamadas bias, en inglés, como parte de cualquiera de las capas ocultas y de la capa de salida.

Estas unidades presentan constantemente un nivel de activación de valor 1. Además, esta unidad está conectada a todas las unidades de la capa superior y los pesos asociados a dichas conexiones son ajustables en el proceso de entrenamiento. El uso de ésta unidad tiene un doble objetivo; mejorar las propiedades de convergencia de la red y ofrecer un nuevo efecto umbral sobre la unidad que opera.

3.3.1 Propagación hacia atrás

Una vez finalizada la fase de propagación hacia adelante se inicia la fase de corrección o fase de propagación hacia atrás. Los cálculos de las modificaciones de todos los pesos de las conexiones empiezan por la capa de salida y continúa hacia atrás a través de todas las capas de la red hasta la capa de entrada. Dentro de los tipos de ajustes de pesos se pueden clasificar dos grupos; ajuste de unidades procesadoras de la capa de salida y ajuste de unidades procesadoras de las capas ocultas.

Ajuste de Pesos de la Capa de Salida: El ajuste de estos pesos es relativamente sencillo, debido a que existe y se conoce el valor deseado para cada una de las unidades de la capa de salida. Cada unidad de la capa de salida produce un número real como salida y se compara con el valor deseado especificado en el patrón del conjunto de entrenamiento. A partir del resultado de la comparación se calcula un valor de error δ_j , según la siguiente ecuación para cada unidad de la capa de salida, siendo t_j el valor de salida deseado y f la derivada de la función umbral f :

$$\delta_j = (t_j - a_j) f'(S_j), \quad (7)$$

Ajuste de Pesos de las Capas Ocultas: Éstas capas no tienen un vector de salidas deseadas y por tanto no se puede seguir el método de propagación de error mencionado en el caso de las unidades procesadoras de la capa de salida. El valor de error calculado para este tipo de unidades procesadoras se obtiene a partir de la siguiente ecuación:

$$\delta_j = [\sum_k \delta_k w_{kj}] f'(S_j), \quad (8)$$

El ajuste de los pesos asociados a las conexiones se realiza a partir del valor generalizado de cada unidad de proceso. Cada peso es ajustado según la siguiente ecuación conocida como la regla generalizada:

$$\delta w_{ji} = \eta \delta_j a_i. \quad (9)$$

La variable η es el coeficiente de aprendizaje. Este coeficiente, normalmente oscila entre 0.25 y 0.75, mismo que refleja el grado de aprendizaje de la red.

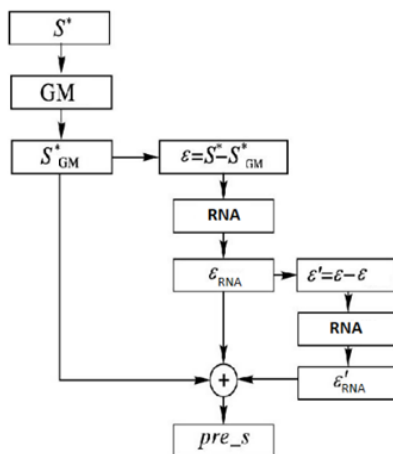


Fig. 5. Esquema del algoritmo predictivo GM.

Algoritmo 1: Algoritmo de la red neuronal Backpropagation

```

1: Algoritmo BP (red datos, η) {
2:   {wij} ← Inicializar;
3:   Mientras ¬ Convergencia(red) Hacer {
4:     e ← seleccionarDatos(datos);
5:     {yk} ← Forward(e);
6:     {dk} ← salidasDeseadas(e);
7:     Para cada nk ∈ CAPA (red, k) Hacer {
8:       δk = (dk - yk) f' (netk);
9:     }
10:    Para j = k - 1 Hasta 1 Hacer {
11:      Para nj ∈ Capa (red, j) Hacer {
12:        δj = f' (netj) ∑i=j+1 δi + Iwj(j+1);
13:      }
14:    }
15:    Para j = k hasta 1 Hacer {
16:      w(j-1)j = w(k-1)j + η δjy(j-1);
17:    }
18:    red ← ActualizarRed({wij});
19:  }
20:  Devolver red;
21: }
  
```

En el algoritmo 1 se aplica el algoritmo para cada conjunto de entrenamiento (línea 4). Se realizan las iteraciones necesarias hasta que el error baje de un umbral determinado. En la siguiente fase de propagación hacia adelante se obtiene el cálculo de la salida de la red (y_k) en la línea 5, y después se obtiene el cálculo de las salidas deseadas de la última capa (línea 6). En la siguiente etapa comienza la propagación hacia atrás.

Primero se obtiene el cálculo de los δ de la capa en función de la siguiente capa (línea 11). Finalmente se lleva a cabo la actualización de pesos de todas las capas (línea 16).

3.4. Algoritmo híbrido propuesto para la previsión de ventas en una tienda de abarrotes

Este artículo propone que el algoritmo híbrido se compone principalmente del análisis, tratamiento y predicción de datos. El tratamiento de datos incluye la detección y eliminación de valores atípicos de datos, la interpolación, reemplazo de valores perdidos y la normalización de datos. Como el esquema del modelo presentado en la figura 4.

3.5. Procesamiento de datos

El pronóstico de ventas se ve afectado por muchos factores. Los datos originales utilizados en el pronóstico tienen ruido de datos, es decir, hay muchos datos extremos o anormales, y el ruido de datos provocará una gran desviación en el pronóstico final. El pre-procesamiento de datos puede hacer que el algoritmo de aprendizaje supervisado obtenga buenos resultados de predicción.

Por lo tanto, en la investigación de este artículo, los datos serán procesados para reducir el ruido de los datos, con el fin de obtener buenos resultados de predicción. El día i proporcionó aquí los datos de ventas expresados como $S_i (i = 1, 2, \dots, n)$, donde n representa el número total de días.

3.6. Detectar y eliminar datos atípicos

Los datos atípicos son datos que se desvían de la mayoría de los datos del conjunto de datos. En el proceso de predicción de datos, los valores atípicos deben eliminarse y tratarse como valores perdidos. Basado en el análisis de muchos datos históricos de ventas, este artículo cree que cuando el conjunto de datos de ventas es $S = \{S_1, S_2, \dots, S_n\}$, S_i satisface las condiciones:

$$abs(S_1 - media(S)) > n * estándar(S). \quad (10)$$

Es decir, encuentra que los datos de ventas S_1 son datos anormales. Entre ellos: $media(S)$ representa el valor promedio de la secuencia S ; $estándar(S)$ representa la desviación estándar de la secuencia S ; $abs(\cdot)$ representa la operación de valor absoluto.

3.7. Interpolación de datos faltantes

En la mayoría de las fuentes de datos originales reales, inevitablemente habrá datos faltantes, y los datos faltantes deben interpolarse para garantizar la integridad de los datos. En este artículo, el valor promedio de los cuatro elementos de datos adyacentes antes y después de los datos faltantes se utiliza para reemplazar el valor perdido.

3.8. Estandarización de datos

La estandarización de los datos puede aumentar la velocidad del cálculo de los datos y, mediante la ésta, los datos con diferentes dimensiones se pueden cambiar a la misma dimensión. En este documento, el método de estandarización de puntaje z se utiliza

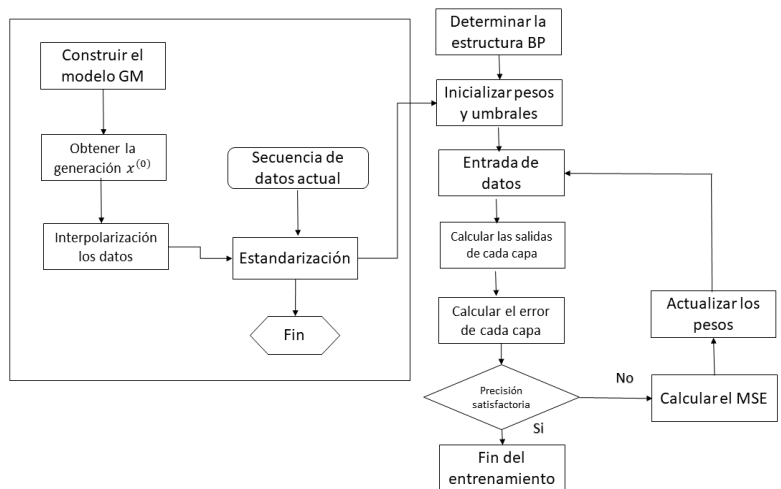


Fig. 6. Diagrama de flujo del algoritmo final. En él se muestra el funcionamiento de la fusión de la red neuronal con el modelo predictivo gris.

Tabla 1. Comparación del rendimiento de cada algoritmo en dos pruebas extras

Modelo	MSE Prueba 1	MSE Prueba 2
ARIMA	14.6 %	13.1%
Serie de tiempo	19.9%	17.4%
GM estadístico (generalizado)	13.0%	10.5%
Red neuronal DGM	11.5%	9.7%

para estandarizar la entrada y la salida. El proceso de estandarización se expresa de la siguiente manera:

$$S_j = media(S) + S'_j .estandar(S). \tag{11}$$

3.9 Predicción del modelo propuesto

Son muchos los factores que inciden en las ventas. Para simplificar el modelo y predecir de forma rápida y eficaz, por lo tanto, estos factores deben ser analizados, a fin de distinguir entre los factores de influencia principales y los factores de influencia secundarios. En los métodos de análisis estadístico matemático, se pueden utilizar muchos métodos de análisis para esto, como el análisis de regresión, sin embargo, estos métodos tienen las siguientes deficiencias:

- La muestra debe obedecer a una determinada distribución de probabilidad típica, existe una relación lineal entre los datos de cada factor y el sistema, y los factores son independientes entre si.

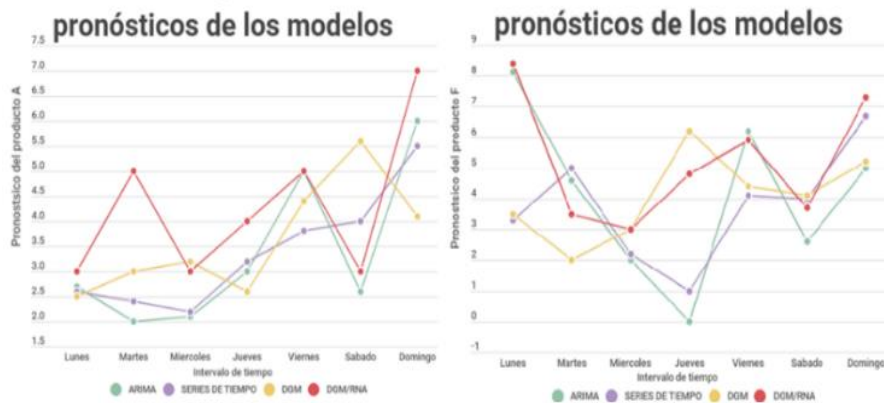


Fig. 7. Gráficas comparativas de los pronósticos obtenidos.

- El análisis requiere una gran cantidad de datos como base y la cantidad de cálculo es grande.
- Puede haber discrepancias entre los resultados cualitativos y cuantitativos.

La literatura [10] y otros modelos que combinan algoritmos lineales y no lineales utilizan principalmente algoritmos lineales para la predicción principal y algoritmos no lineales para la predicción residual. La fórmula de cálculo es la siguiente:

$$S^* = S^{*lineal} + (S^* - S^{*lineal})_{no\ lineal} + \varepsilon_1, \tag{12}$$

$$\hat{S}^* = S^{*lineal} + (S^* - S^{*lineal})_{no\ lineal}, \tag{13}$$

dónde: S^* es la secuencia de ventas procesada; $S^{*lineal}$ es el resultado de la predicción del algoritmo lineal; $(S^* - S^{*lineal})_{no\ lineal}$ es el resultado del cálculo del algoritmo no lineal; $(S^* - S^{*lineal})_{no\ lineal}$ como ε de la fórmula (14) 1 factores se añadió e influye en los resultados de las ventas, las ventas de predicción no lineal; S^* es el resultado de la predicción de todo el modelo predictivo y ε_1 es el residuo de predicción de todo el algoritmo.

En la figura 5, ε_{RNA} representa el resultado de usar el algoritmo RNA para predecir ε ; ε' es el residuo de la predicción RNA de ε . El resultado final de la predicción de la prueba experimental es:

$$pre_s = S^{*GM} + \varepsilon_{RNA} + \varepsilon'_{RNA}, \tag{14}$$

dónde ε'_{RNA} es el resultado de la predicción del ε residual que la RNA predice nuevamente. En el proceso de pronóstico real, el modelo gris se usa principalmente para predecir datos históricos de ventas: el primer paso la red neuronal se usa para analizar los residuos de ventas pronosticados por el modelo gris, luego, los factores que tienen un mayor impacto sobre las ventas se suman a la predicción. Posteriormente la RNA realiza la predicción sobre el residual entre el volumen de ventas real y el resultado de la predicción como está ilustrado de la figura 6.

3.10 Cálculo del MSE

MSE básicamente mide el error cuadrado promedio de las predicciones. Para cada punto se calcula la diferencia cuadrada entre las predicciones y el objetivo y luego promedia esos valores. Para ello, se suman todos los errores y se dividen entre el número total de puntos. N representa el número total de puntos dando lugar a la fórmula del Error Cuadrático Medio. Dicho cálculo se denota por la siguiente ecuación:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i), \quad (15)$$

donde y_i es el resultado real esperado y \hat{y}_i es la predicción del modelo.

4. Resultados experimentales

En este artículo, el proceso de pronóstico se utilizó para predecir los datos de ventas, combinado con las características de datos de ventas limitados, gran ruido de datos y grandes fluctuaciones de datos, el algoritmo GM y el modelo de la RNA es utilizado para realizar predicciones de datos de ventas de productos.

Dentro de la hibridación de estos algoritmos, el modelo gris puede obtener resultados de predicción mejores y estables cuando la cantidad de datos es limitada, por otra parte, la red neuronal tiene un buen rendimiento en la predicción de secuencias no lineales. Los datos originales utilizados en la investigación de este artículo provienen de los datos de ventas reales de productos de una tienda de abarrotes.

Se puede ver en la Tabla 1 que el porcentaje del MSE generado por el algoritmo híbrido propuesto es de aproximadamente 11.5% cuando se utiliza una muestra pequeña del conjunto de datos para la predicción. El rendimiento del algoritmo es mejor que los otros modelos empleados por el cliente, debido a que estos dos últimos requieren un mayor volumen de datos para su análisis, mientras que el algoritmo híbrido es más estable y la precisión de la predicción es relativamente alta; se puede ver que la precisión y la estabilidad de la predicción del algoritmo propuesto mejora a comparación de los otros tres algoritmos.

De igual manera, el pronóstico fue satisfactorio al tener un MSE relativamente bajo como se muestra en la figura 7. La predicción residual secundaria del algoritmo inteligente híbrido puede ser muy eficaz, mejora la precisión y estabilidad de los resultados finales del pronóstico en gran medida y brinda altas perspectivas de su aplicación en los pronósticos de ventas reales.

5. Conclusiones y trabajo a futuro

En un pequeño comercio de abarrotes (pymes), la previsión de ventas es una parte importante de la gestión del suministro de productos y minoristas, lo que tiene un gran impacto en la mejora del beneficio final de la pyme. Aunque ya se emplean muchos modelos para hacer previsiones de ventas de múltiples productos, la cantidad limitada de datos y las grandes fluctuaciones en los datos generales son siempre los factores que limitan la aplicación de estos modelos a las previsiones de ventas totales.

Este artículo propone un algoritmo híbrido, que combina el algoritmo GM y una red neuronal. Incluye principalmente dos partes: tratamiento de datos y predicción de datos. En el proceso de predicción de datos, el algoritmo GM se utiliza para predecir la secuencia mientras que la RNA predice el residuo de la predicción del modelo gris y posteriormente agrega factores que tienen un gran impacto en las ventas.

En el proceso para la obtención del pronóstico, debido a las fluctuaciones relativamente grandes en las ventas, se deben considerar múltiples opciones que brinden un pronóstico certero, o que por lo menos, tengan un mínimo margen de error. A través del pronóstico de dos productos de la pyme y las ventas de las tiendas, se concluye que el algoritmo inteligente híbrido propuesto en esta investigación tiene un margen de error relativamente pequeño, pero cuenta con una mayor precisión, a pesar de haber empleado una muestra limitada de datos. La estabilidad de predicción del algoritmo en este documento es alta y se puede generalizar su aplicación a varios sectores.

También se debe mencionar que, debido a la pandemia, hoy en día es más costoso y complicado obtener un pronóstico muy certero, debido a la volatilidad y anomalías presentes por la reciente situación.

Referencias

1. Adya, A. M., Collopy, F.: How effective are neural networks at forecasting and prediction? A review and evaluation. *Journal of Forecasting*, vol. 17, no. 5-6, pp. 481–495 (1998)
2. Chase, C.: Ways to improve sales forecasts. *The Journal of Business Forecasting Methods & Systems*, tomo 12, no. 3, pp. 15–17 (1993)
3. Stubbings, P., Virginas, B., Owusu, G., Voudouris, C.: Modular neural networks for recursive collaborative forecasting in the service chain. *Knowledge-Based Systems*, vol. 21, no. 6, pp. 450-457 (2008)
4. Deng, J. L.: *Grey system theory*, Publishing House of Huazhong University of Science and Technology, Wuhan, China (2002)
5. Li, S., Cheng, Y.: Deterministic fuzzy time series model for forecasting enrollments. *Computers and Mathematics with Applications*, vol. 53, no. 12, pp. 1904–19020 (2007)
6. Zhang, G.: Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, vol. 50, pp. 159–175 (2003) doi: 10.1016/S0925-2312(01)00702-0
7. Wong, W. K., Guo, Z.: a hybrid intelligent model for medium-term sales forecasting in fashion retail supply chains using extreme learning machine and harmony search algorithm. *International Journal of Production Economics*, vol. 128, no. 2, pp. 614–624 (2010) doi: 10.1016/j.ijpe.2010.07.008
8. Hsu, C. C., Chen, C. Y.: Applications of improved grey prediction model for power demand forecasting. *Energy Conversion and Management*, vol. 44, no. 14, pp. 2241-2249 (2003)
9. Xie, N., M., Liu, S. F.: Discrete GM 1, 1 and mechanism of grey forecasting model. *Xitong Gongcheng Lilun yu Shijian, System Engineering Theory and Practice*, vol. 25, no. 1, pp. 93–99 (2005)
10. Zhang, D., Zhang, X.: Study on forecasting the stock market trend based on stochastic analysis method. *International Journal of Business and Management*, vol. 4, no. 6, pp. 163–170 (2009)
11. Xie, N. M., Liu, S. F.: Research on discrete grey model and its predictive result. *Journal Systemy Enregy*, pp 520–523 (2006)

12. Xie, N. M., Liu, S. F.: Discrete grey forecasting model and its optimization. *Applied Mathematical Modelling*, vol. 33, no. 2, pp. 1173–1186 (2009) doi: 10.1016/j.apm.2008.01.011
13. Li, D. C., Chang, C. J., Chen, W. C., Chen, C. C.: An extended grey forecasting model for omnidirectional forecasting considering data gap difference. *Applied Mathematical Modelling*, vol. 35, no. 10, pp. 5051–5058 (2011) doi: 10.1016/j.apm.2011.04.006
14. Lactermacher, G., Fuller, J. D.: Backpropagation in time-series forecasting. *Journal of Forecasting*, vol. 14, no. 4, pp. 381–393 (1995) doi: 10.1002/for.3980140405