

Contexto

En esta presentación se comparan los papers proporcionados, que tratan de los modelos de RNN LSTM y GRU.

Metodología de búsqueda

Texto encontrado en Google Scholar. Traté de buscar algún paper que incluyera LSTM y GRU.

Términos Preliminares

- **NN:** Neural Networks.
- **RNN (Recurrent Neural Network):** tipo de red neuronal artificial que utiliza datos secuenciales o datos de series temporales. Estos algoritmos de deep learning se usan comúnmente para problemas ordinarios o temporales como traducción de idiomas o procesamiento del lenguaje natural (NLP). A diferencia de las *Feedforward* NN, que procesan datos en una sola pasada, las RNN procesan datos en múltiples pasos de tiempo.
- **Vanishing/Exploding gradient Problem:** Se produce cuando los gradientes de pérdida de la red con respecto a los parámetros (weights) se vuelven excesivamente grandes o demasiado pequeños. Puede provocar inestabilidad numérica y la incapacidad de la red para converger a una solución adecuada.
- **MAE:** Mean Absolute Error.
- **MSE:** Mean Square Error.
- **RMSE:** Root Mean Square Error.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad RMSE = \sqrt{\sum_{i=1}^n (P_i - A_i)^2 / n}$$

Términos Preliminares

- **LSTM (Long Short-Term Memory):** Un tipo de RNN diseñado para evitar los problemas de vanishing y exploding gradient problems.
- **BPTT:** Back-Propagation Through Time.
- **RTRL:** Real-Time Recurrent Learning.
- **ARIMA:** autoregressive integrated moving average. Predice valores futuros basándose en valores pasados. ARIMA utiliza promedios móviles rezagados para suavizar los datos de series temporales.
- **CEF (Constant Error Flow: Incluido en LSTM):** está diseñado para abordar el problema del desvanecimiento del gradiente en las RNN tradicionales, permitiendo que la información fluya a través de la red de manera más estable durante el entrenamiento.
- **CEC (Constant Error Carrousel):** Utilizado para manejar el problema del desvanecimiento del gradiente en redes neuronales que trabajan con secuencias largas.

*CEF en las LSTM se centra en mantener un flujo constante de información a través de una celda de estado, mientras que el CEC es una estrategia similar en el sentido de mantener la información a través de la red, pero el enfoque y los detalles pueden variar. Ambos métodos tienen el objetivo común de preservar información a lo largo de secuencias y mejorar la capacidad de la red para aprender dependencias a largo plazo.

Week 6 Research Stay

Comparación [1]

En esta ocasión, me centraré en realizar una comparativa entre el último texto proporcionado, "Using LSTM and GRU neural network methods for traffic flow prediction" (Ref. 3) y el que encontré, "Peak Electrical Energy Consumption Prediction by ARIMA, LSTM, GRU, ARIMA-LSTM and ARIMA-GRU Approaches" (Ref. 4) debido a que los otros dos papers son solo introducciones a las RNN LSTM y GRU respectivamente y no da mucho para generar esta comparativa.

En Ref. 3, el objetivo fue demostrar que, después de un experimento realizado con ARIMA, LSMT y GRU para predecir el tráfico vehicular, estos dos últimos modelos dieron resultados más acertados que los predichos con ARIMA, que es bastante más viejo en comparación. Según menciona, "En promedio, las NN GRU han reducido el MAE a un nivel de aproximadamente 10% que el modelo ARIMA y 5% que el modelo NN LSTM." Los resultados de este estudio se muestran en la siguiente gráfica

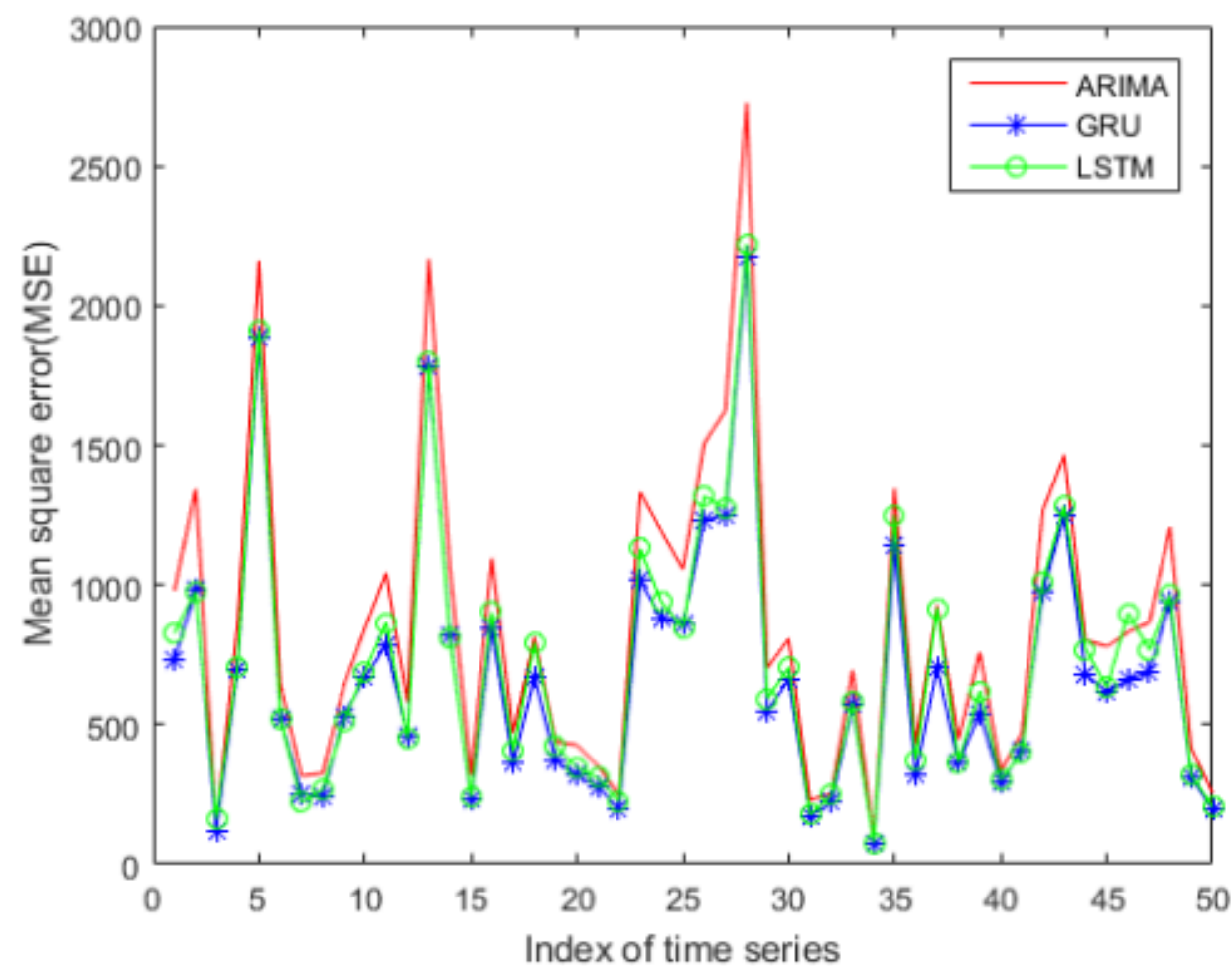


Fig. 3. MSE over 50 selected time series.

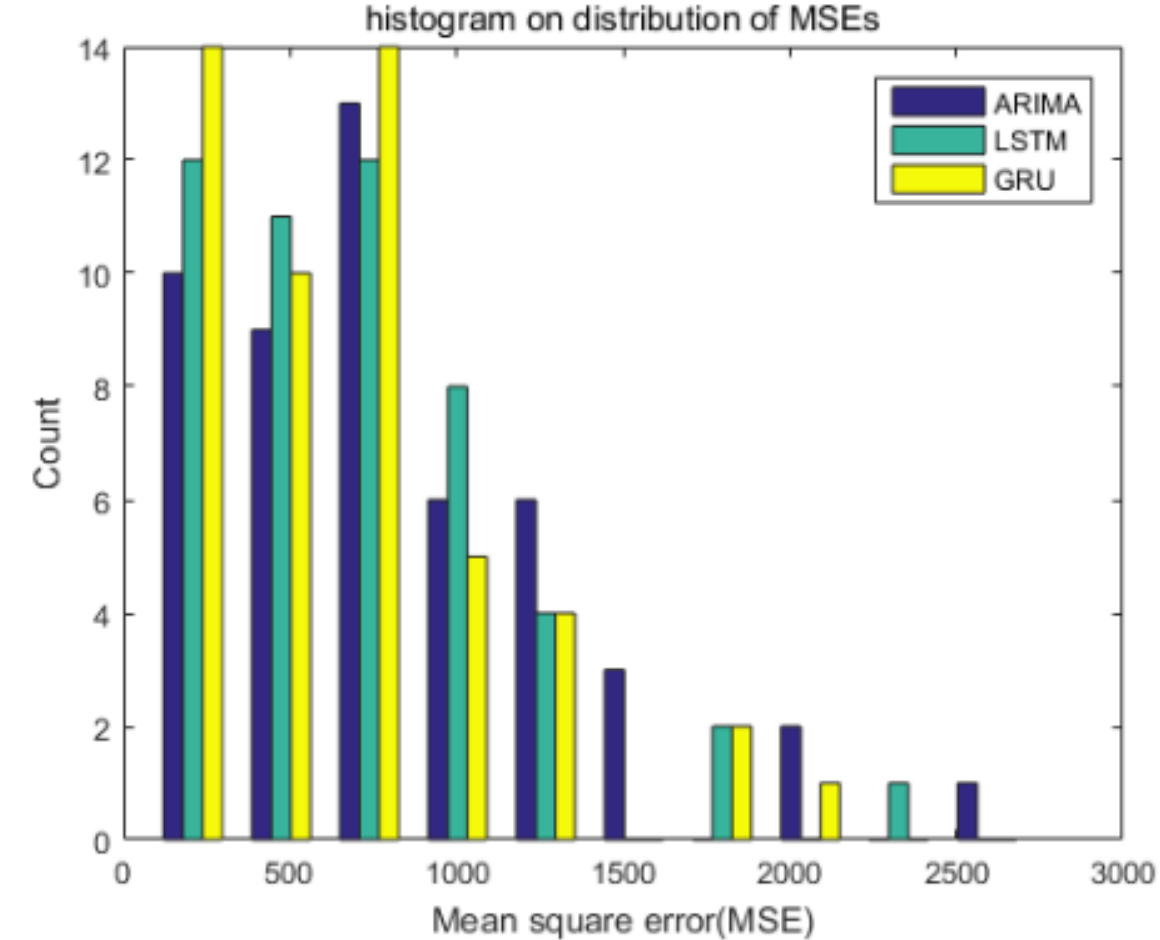
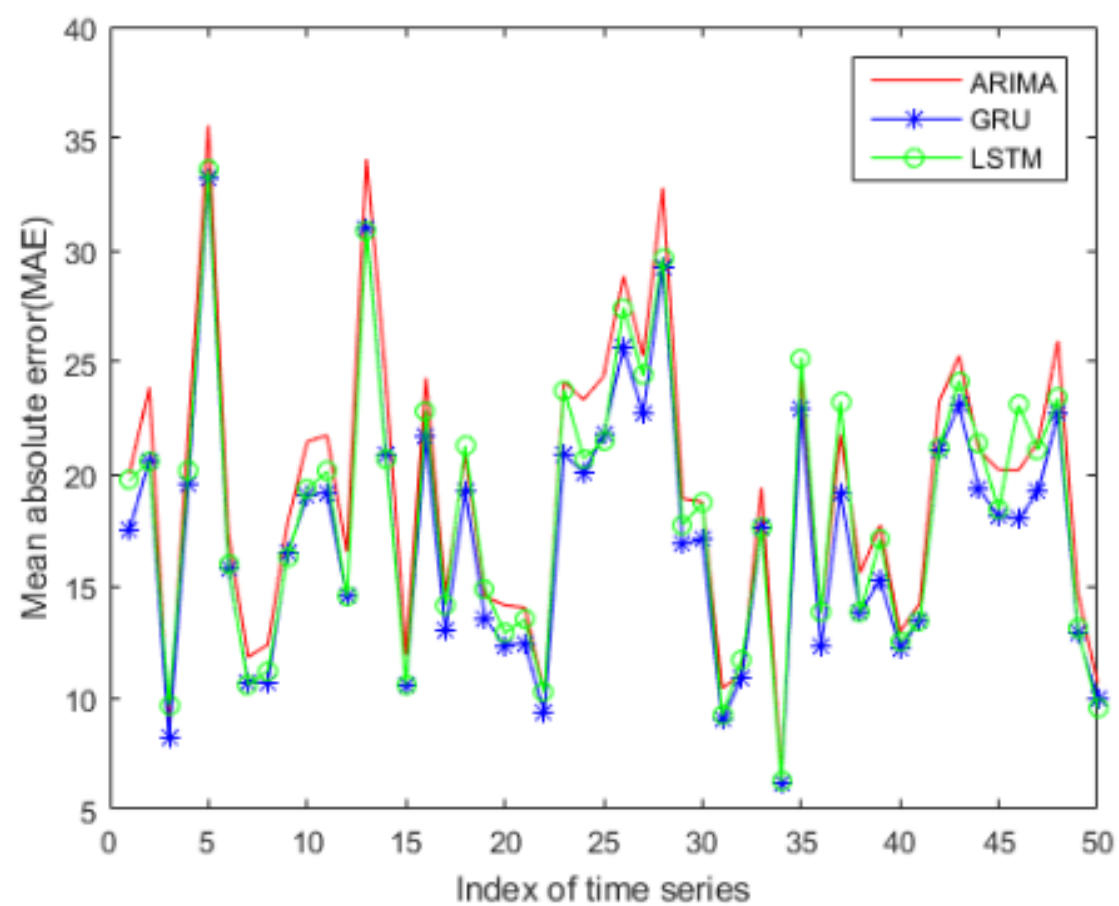


Fig. 5. Histogram on distribution of MSEs

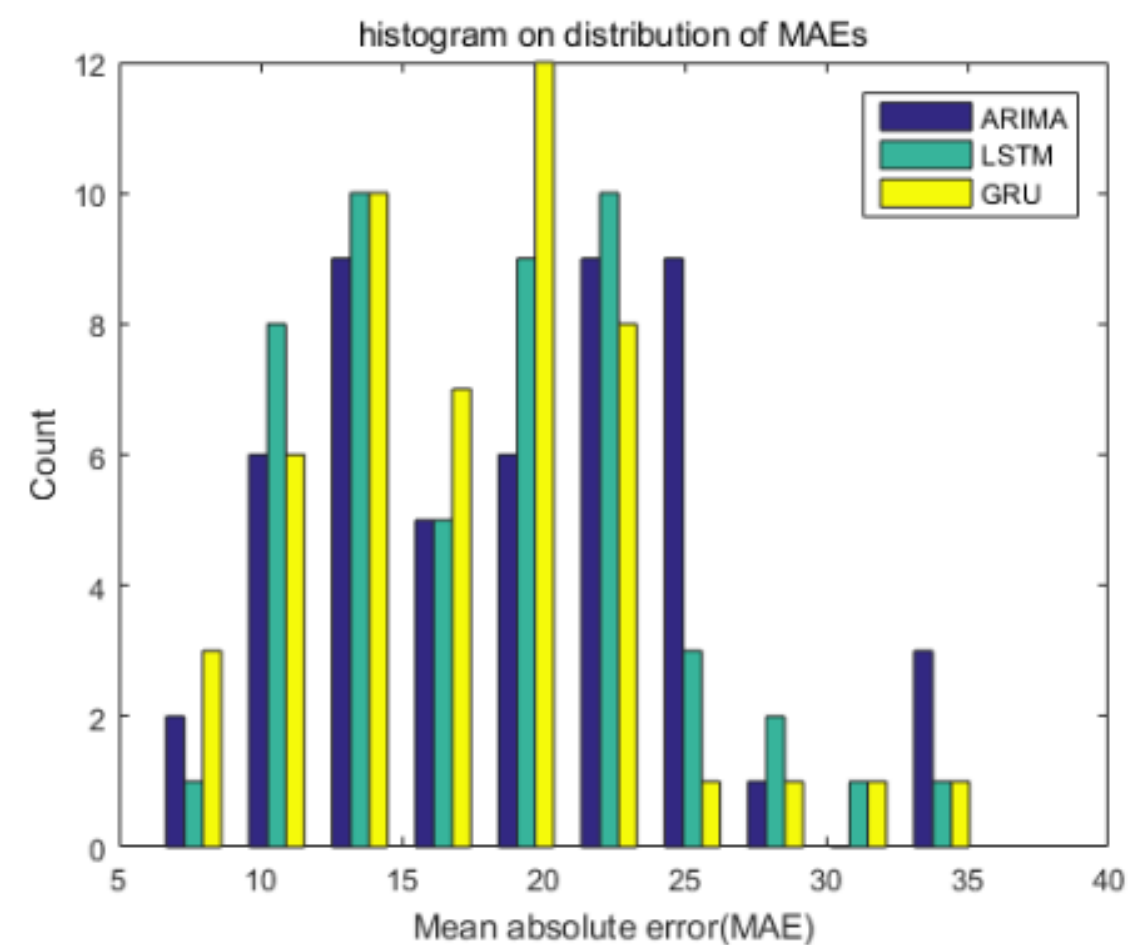


Fig. 6. Histogram on distribution of MAEs.

Comparación [2]

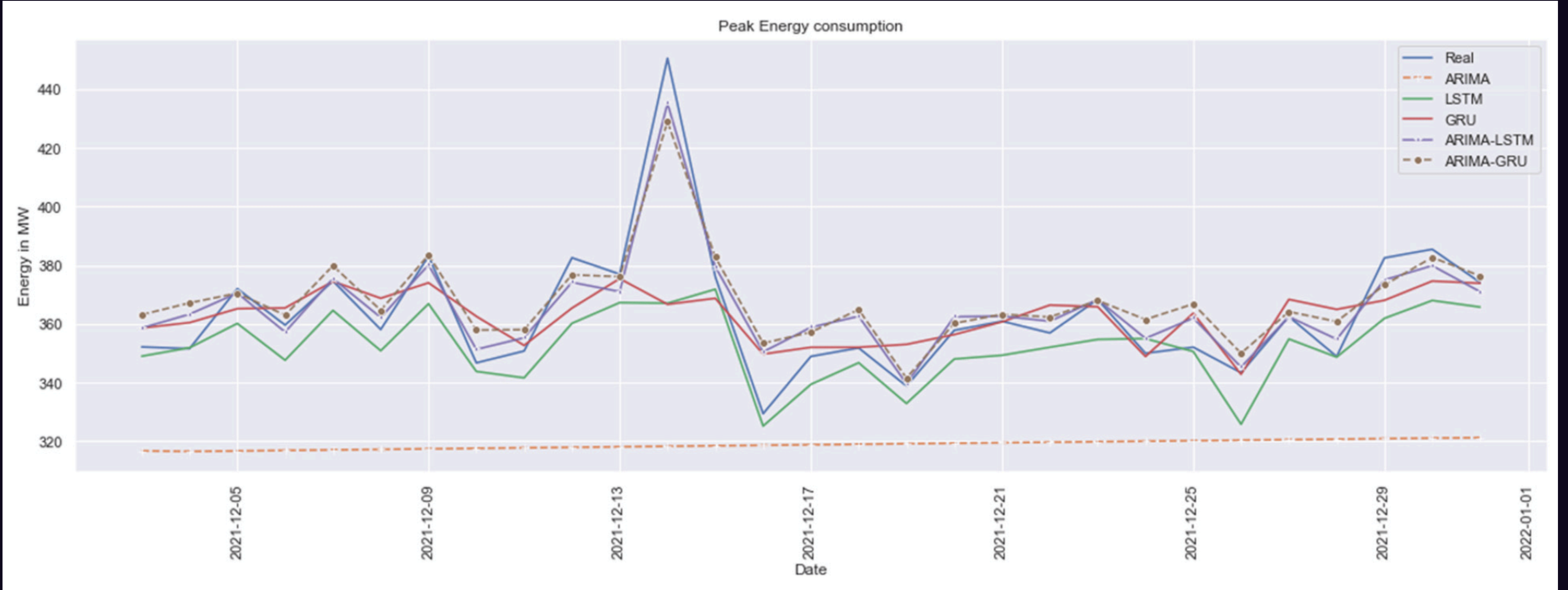
Parecido a este experimento, se realizó algo parecido pero ahora para tratar de predecir el punto más alto del consumo de electricidad. Además, a comparación de los métodos utilizados en el paper anterior (Ref. 3) (hecho con ARIMA, LSTM y GRU), éste lo realiza, además de con los tres anteriores, con enfoques híbridos ARIMA-LSTM y ARIMA-GRU.

El consumo pico de energía eléctrica se predijo utilizando enfoques híbridos basados en métodos tradicionales de predicción de series de tiempo (ARIMA) y métodos de aprendizaje profundo LSTM y GRU. El enfoque ARIMA se utilizó para modelar el término de tendencia, mientras que se emplearon enfoques de aprendizaje profundo para interpretar el término de fluctuación, y los resultados de estos modelos se combinaron para proporcionar el resultado final.

Según este paper, el enfoque híbrido ARIMA-LSTM proporcionó el mejor rendimiento de predicción con un RMSE de 7,35, mientras que para el enfoque híbrido ARIMA-GRU, el RMSE fue de 9,60. En general, los enfoques híbridos superaron a los enfoques individuales, como GRU, LSTM y ARIMA, que exhibieron valores de RMSE de 18,11, 18,74 y 49,90, respectivamente.

Table 2. Statistical indicators.

Approaches	ARIMA	LSTM	GRU	ARIMA-LSTM	ARIMA-GRU
RMSE	49.90	18.74	18.11	7.35	9.60
MAPE	12.05	3.01	2.60	1.52	1.56



Referencias

- [Ref. 1] Long Short-Term Memory.
- [Ref. 2] Tutorial on GRU
- [Ref. 3] Using LSTM and GRU neural network methods for traffic flow prediction
- [Ref. 4] Peak Electrical Energy Consumption Prediction by ARIMA, LSTM, GRU, ARIMA-LSTM and ARIMA-GRU Approaches