Predicción de la velocidad de circulación a utilizando una red neuronal LSTM y GRU.

Gabriel Díaz Calderón  
Departamento de Informática  
Universidad Andrés Bello Santiago, Providencia  
[g.diazcalderon@uandresbello.edu](mailto:g.diazcalderon@uandresbello.edu) Daniel Montecinos Zambra  
Departamento de Informática  
Universidad Andrés Bello Santiago, Providencia  
[d.montecinosz@uandresbello.edu](mailto:d.montecinosz@uandresbello.edu)

**Términos generales**

**Predicción, LSTM**

**Palabras claves**

**Predicción, velocidad, red, LSTM, GRU**

1. Introducción

Las redes neuronales recurrentes con memoria a largo plazo (LSTM) han surgido como un modelo eficaz y escalable para varios problemas de aprendizaje relacionados con datos secuenciales.

Este tipo de red con el tiempo se ha ido estableciendo y practicando para resolver un variado tipo de problemas. Esto incluye el reconocimiento de escritura [4] - [5] y la generación [6], el modelado del lenguaje [7] y la traducción [8], el modelado acústico del habla [9], la síntesis del habla [10], la predicción de la estructura secundaria de la proteína [11], Análisis de audio [12], y datos de video [13] por citar algunos.

El siguiente documento describe la investigación realizada para predicción de velocidad a partir de datos temporales de camiones del centro de transporte y logística de la Universidad Andrés Bello, Chile. Estos datos son obtenidos desde el monitoreo GPS satelitales.

El presente documento está estructurado de la siguiente forma, la sección dos presenta el trabajo relacionado. La sección dos presenta los trabajos relacionados a nuestra investigación. La sección 3 presenta el análisis de datos. La sección 4 presenta la propuesta realizada. La sección 5 presenta sus resultados.

2. Trabajos relacionados

En general, realizar una predicción de datos no es una tarea sencilla. Para la investigación, se utilizaron como referencia un estudio realizado en 2017 el cual propone un nuevo modelo de pronóstico de tráfico basado en una red de memoria a largo plazo a largo plazo (LSTM) [2] utilizando datos de Beijín del 2015 con un intervalo de 5 minutos y una cantidad total de 25 millones de datos. En este estudio se propone la utilización de la red LSTM tomando en cuenta el carril por el cual transitan los vehículos y la velocidad media. El resultado de su experimentación es de MRE 6.01% atribuido a la media de errores. Otro trabajo que considerado es el Uso de los métodos de redes neuronales LSTM y GRU [3] para la predicción del flujo de tráfico. En esta investigación se comparan dos redes neuronales recurrentes diferentes para la predicción del tráfico utilizando 50 sensores con datos cada 30 segundos de diferencia. Los resultados fueron favorables a la utilización de la red neuronal GRU con una media de valor absoluto de 17.2

3. Data

Cada conjunto de datos a presentar se divide en dos partes: un conjunto de entrenamiento, y un conjunto de pruebas para la evaluación final en la red LSTM.

3.1 Data de transporte

Se utilizaron datos del Centro de Transporte y Logística de la Universidad Andrés Bello, Chile; obtenidos de sus monitoreos GPS satelitales en la ciudad de Santiago. Esta recolección de datos GPS, en algunas ocasiones presentan problemas por interferencias, como se ilustra en la fig.1, debido a grandes alturas, alta nubosidad, túneles, por citar algunos. Esto consideró en el procesamiento de datos que se puede ver en la sección 3.3.

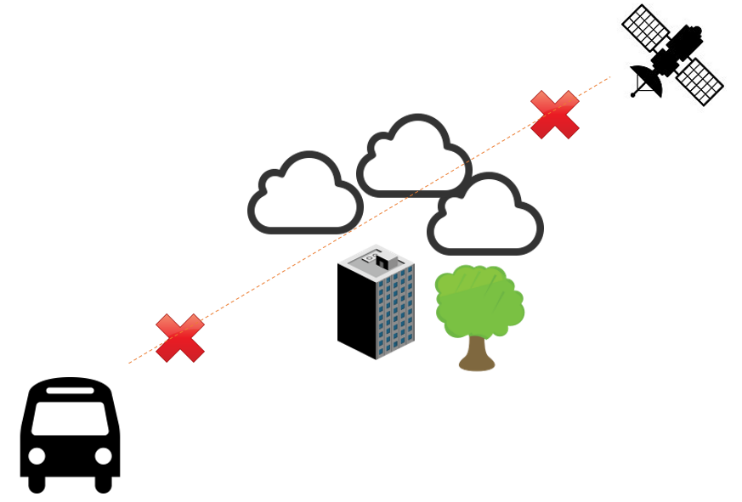


Fig.1 – Interferencias datos de transporte

Los datos están agrupados en cuadrantes de 500x500 del año 2016. Se cuentan con 10.9 Millones de datos disponibles. Los datos cuentan con 12 features presentes. Los cuadrantes presentes en la data son 5735, con los cuales se trabaja en la red. Se pueden visualizar los features en la siguiente tabla 1.

|  |  |
| --- | --- |
| Numero | Feature |
| 1 | Id\_Detencion |
| 2 | Id\_Calle |
| 3 | Id\_Cuadrante |
| 4 | Nombre\_Comuna |
| 5 | Movil |
| 6 | Fecha |
| 7 | Dia |
| 8 | Hora\_inicio |
| 9 | Hora\_fin |
| 10 | Min\_detenido |
| 11 | Latitud |
| 12 | Longitud |

Tabla 1. Features data transporte

3.2 Datos de humedad

Para la investigación, se utilizaron datos de humedad relativa obtenidos de (pagina), en donde se cuenta con 11 centros de monitoreo en la región metropolitana. Estos centros cuentan con datos históricos los cuales van desde el 2004 hasta la actualidad. En algunos casos, estos datos son despreciados en algún momento temporal, por lo cual se realizó un catastro de los datos disponibles en los centros de monitoreo. Se cuentan con los datos de 2018 de los siguientes centros de monitoreo. Véase tabla 2.

|  |  |  |
| --- | --- | --- |
| Centro de monitoreo | | Humedad |
| Cerillos | Sin datos disponibles en 2016. | |
| Cerro Navia | Data disponible | |
| El Bosque | Data disponible | |
| Independencia | Data disponible | |
| La Florida | Data disponible | |
| Las Condes | Data disponible | |
| Parque O’Higgins | Data disponible | |
| Pudahuel | Data disponible | |
| Puente alto | Data disponible | |
| Quilicura | Datos disponibles hasta 2016 | |
| Talagante | Data disponible | |

Tabla 2 – Data centros de monitoreo RM

4. Procesamiento de datos

Para el preprocesamiento de los datos se consideró como se menciona en el punto 3.1, las potenciales interferencias sufridas en la obtención de datos por el monitorio GPS.

4.1 Missing data Datos transporte

Con ayuda de la herramienta R studio, para los casos expuestos en el siguiente punto, podemos deslumbrar que no disponemos de missing data, contemplando un 0% de missing data.

4.2 Missing data Humedad

En cuanto a la missing data de la humedad relativa, se percató sí se existía (3398 datos faltantes), tal como se muestra en fig.2. La missing data presente se reemplazó por el promedio de la estación.

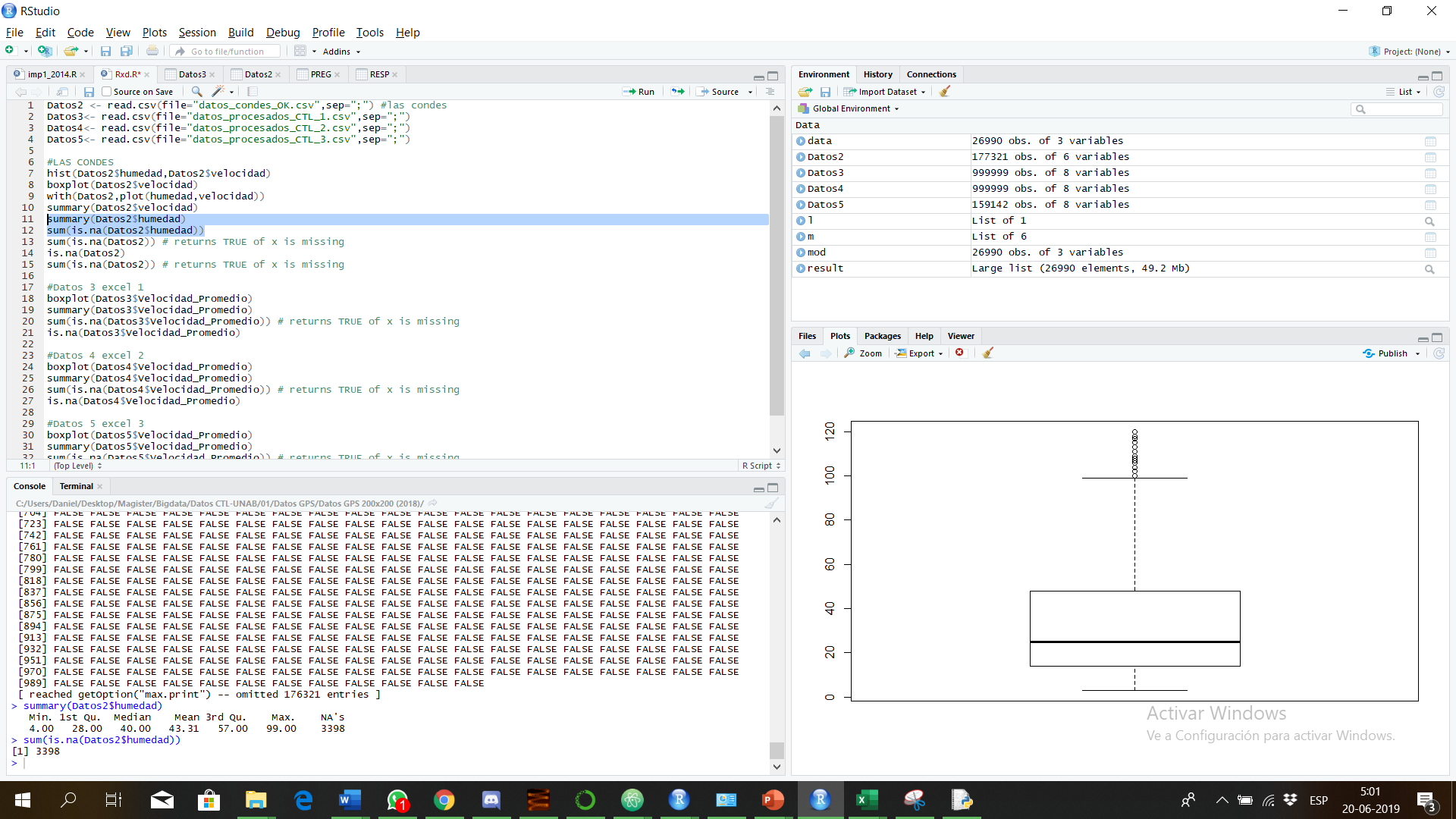


Fig2. Missing data humedad

4.3 Outliers

Los outliers presentes son debido a lo ya comentado en las secciones previas 3.1 y 3.3. La data presente del centro del transporte y logística contaba con la presencia de datos anormales los cuales no podían suceder. Tales sean casos de velocidades excedentes a los 120 km/h, los cuales se trabajaron y se corrigieron. Se realizó una limpieza de estos datos delimitando a 120km/h la velocidad máxima. Con esto, se eliminaron 6573 datos atípicos. Quedando los datos de la siguiente forma más normalizado a un comportamiento común de los datos. En fig 3. Y fig.4 se puede ver el antes y después de la limpieza de los datos.

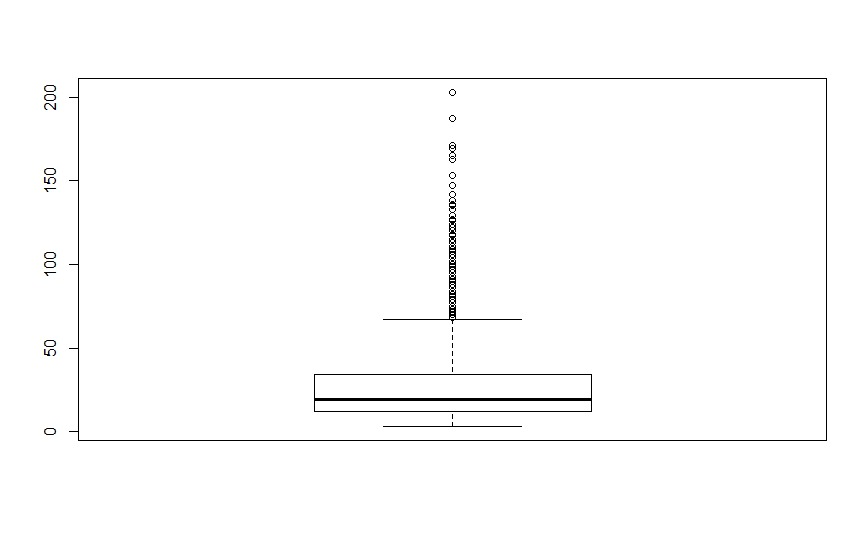


Fig3. Antes de limpieza de outliers.

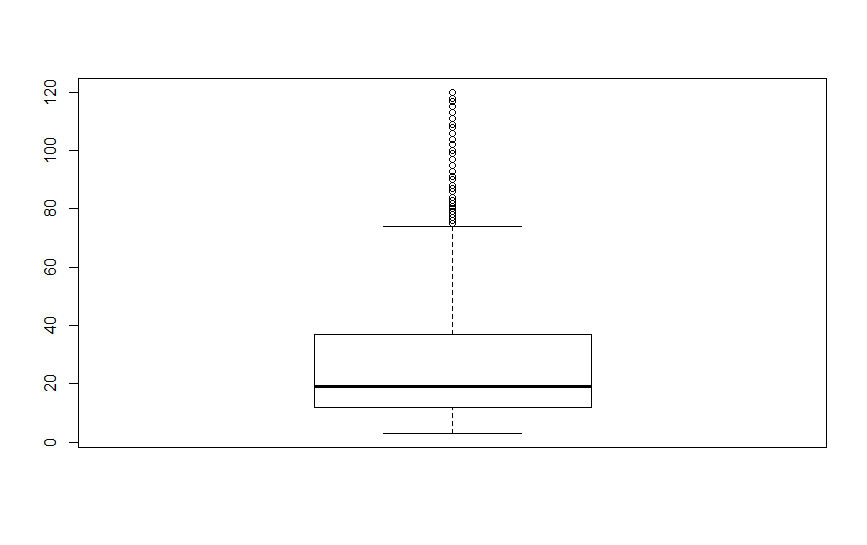


Fig.4 Después de limpieza de outliers.

4.4 Análisis de datos

Para los datos del centro de transporte en adición se hizo un resumen de datos para una previa referencia. Las tablas 3,4 y 5 presentan el análisis de los datos en diferentes agrupaciones, en donde se denota su leve desviación.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Min | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
| 3.00 | 12.00 | 19.00 | 30.47 | 37.00 | 120.00 |

Tabla 3. Analisis de data agrupación 1.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Min | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
| 3.00 | 12.00 | 21.00 | 31.94 | 41.00 | 120.00 |

Tabla 4. Analisis de data agrupación 2.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Min | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
| 3.00 | 14.00 | 25.00 | 35.67 | 48.00 | 120.00 |

Tabla 5. Analisis de data agrupación 3.

4.5 Correlación de datos

La correlación de datos se realizó para establecer la coherencia de dos factores. El primero, entre la velocidad y las velocidades anteriores. El segundo, entre velocidad y humedad, en donde nos damos cuenta de que existe una muy baja correlación. Con esto, se planteará a futuro que hacer respecto a esta situación. Para la correlación de humedades se asume la humedad como lluvia, tal que humedad sea mayor a 90. Se representa gráficamente las correlaciones en la fig.3.

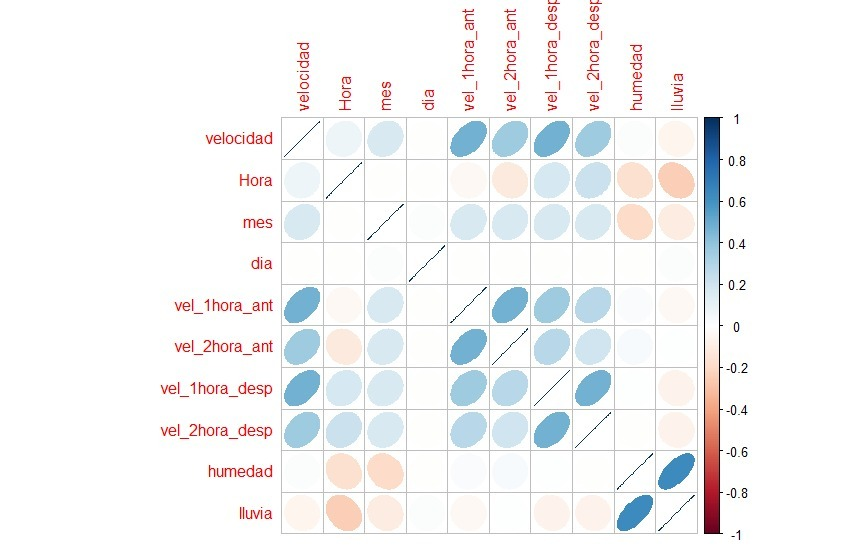


Fig. 3 – Grafico de correlación de datos

4.6 PCA

En análisis de componentes principales se realizó cuantitativamente y gráficamente como se aprecia en fig.4. Y fig.5.

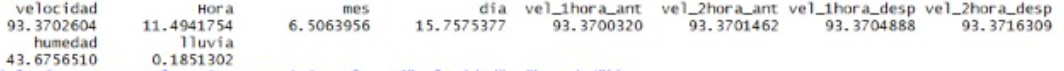


Fig4. Análisis de componentes principales

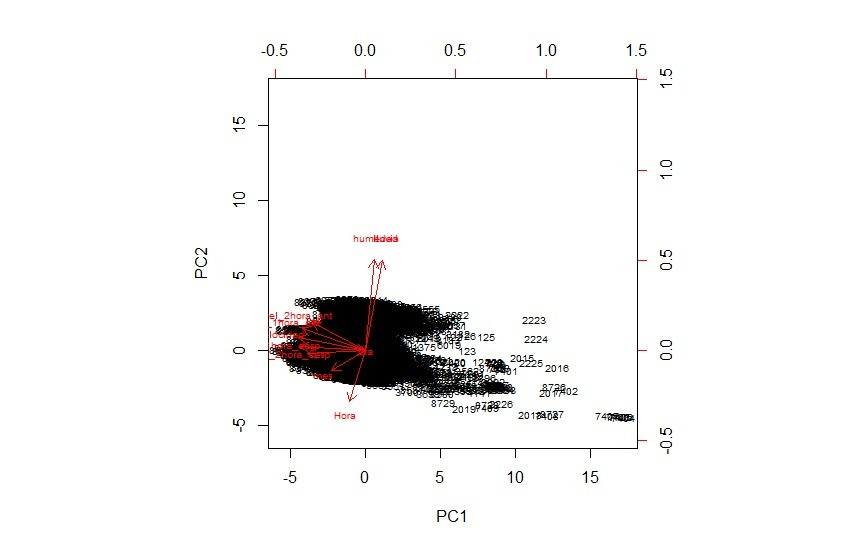


Fig.5 Grafico de análisis de componentes principales.1

# propuesta

La propuesta de esta investigación es integrar los features descritos en la sección 3. Las temporalidades descritas en conjunto de la velocidad integradas en la red recurrente LSTM.

5.1 Red neuronal recurrente LSTM Y GRU

Como propuesta de red neuronal se utilizarán las redes neuronales recurrente Long Short Term Memory (LSTM) y la red Gated Recurrent Units (GRU) como se muestra en la Fig. 5 y Fig. 6 respectivamente. Estas redes nos permitirán trabajar con datos de series de tiempo para el aprendizaje. Para la utilización de la red neuronal utilizaremos Python 3.6, Tensorflow 1.10, Keras 2.2 y los datos de la fecha (transformados a día y mes), la hora, la velocidad de una hora anterior y la humedad del aire, todos estos datos serán escalados entre 0 y 1. A través de estos datos (incluyendo la velocidad de una hora antes) se predecirá la velocidad de una hora después.

La arquitectura de la red se encuentra conformada por una capa de entrada, una capada de salida y una capa de olvido la cual nos permitirá extraer los datos que no son relevantes para el aprendizaje. Como función de costo se utilizará la media del valor absoluto y como función de activación se utilizará la función Sigmoide.

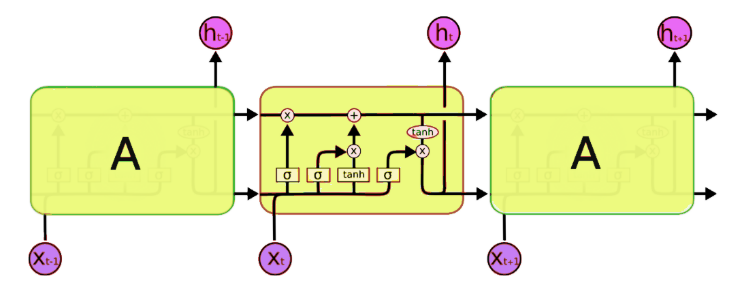


Fig5. Long short team memory LSTM.

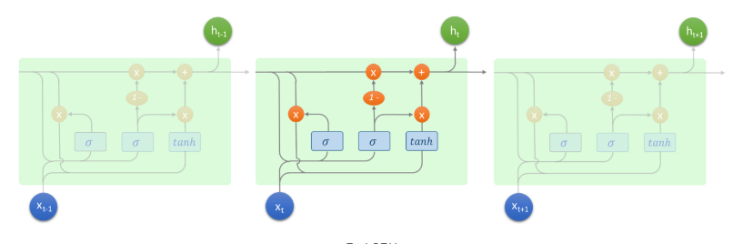


Fig6. Gated recurrent units GRU.

# Procesamiento

Para mejorar el funcionamiento de nuestro algoritmo de Deep Learning se normalizaron los datos comprimiendo los valores de estos. Para esto se utilizó un escalado estándar restando a cada dato la media de la variable y dividiendo por la desviación típica como se muestra en la Fig. 7.

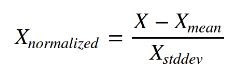
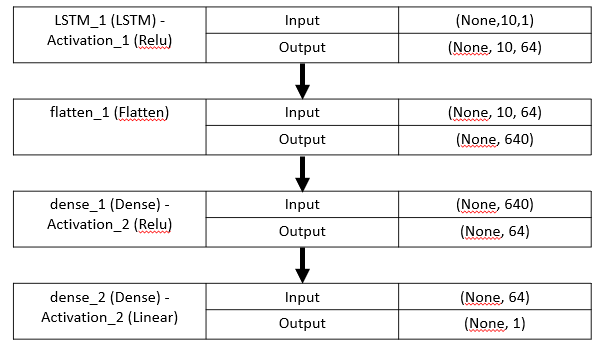


Fig.7 Ecuación de Escalado Estandar.

Para nuestro entrenamiento con series de tiempo se utilizarán *N* horas antes para predecir *M* horas después. Como la información de los primeros *N* y últimos *M* datos no se registra se decidió eliminar estos datos para no generar un ruido innecesario en nuestro set de entrenamiento.

# Arquitectura



|  |  |
| --- | --- |
| Modelo | MSE |
| LSTM | 0.1939 |
| GRU | 0.1041 |

Fig 8. Arquitectura de la red

Como se muestra en la Fig 7. nuestro modelo desarrolla una RNN LSTM la cual tiene 10 datos de entradas equivalentes a las 10 horas anteriores que se usarán para predecir 1 hora siguiente, como función de activación se utiliza la función de unidad lineal rectificada (ReLu) para la disminución del error, para la disminución de dimensiones se utiliza una capa Flatten la que nos permite la utilización de un vector en nuestra siguiente capa, a continuación se utiliza una capa oculta Full-Conection (Dense) con una función de activación ReLu. Como capa de salida se ocupa una capa Dense con una función lineal que nos permitirá la utilización del escalar entre valores positivos y negativos.

# Entrenamiento

Como datos de entrenamiento utilizaremos un solo cuadrante utilizando datos desde el 1 de Enero hasta el 31 de Diciembre del año 2016,

La función de costo para nuestro modelo se probará utilizando 3 funciones de costo; Mean Squred Error (MSE); Mean Absolute Error (MAE); Mean Relative Error (MRE).

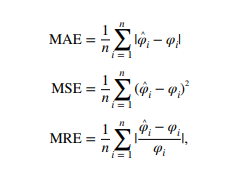


Figura 9. Funciones de Costo

Como se muestra en la Fig. 9 donde φ ^ i son los datos de pronóstico, mientras que φi son los datos medidos. La función MAE y el MSE son más sensibles a los datos de tráfico sin procesar. Por lo tanto, el MRE es más adecuado para servir como criterio de evaluación cuando se compara con otros modelos de pronóstico de tráfico [7].

# 9. resultados

Los resultados obtenidos por parte de la red LSTM se puede apreciar en la fig.10. en donde se logra apreciar con color rojo los datos predichos y con color verde los datos reales, en lo cual se denota un gran aprendizaje. Utilizando la métrica de evaluación MSE se puede diferenciar mínimamente el promedio de errores cuadráticos.

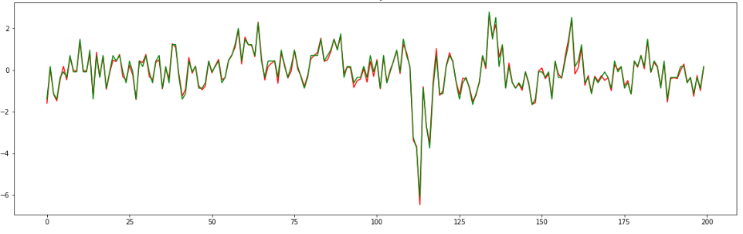


Fig.10 – Resultados

Figura 11. Evaluación de métricas

7. Conclusiones y trabajos futuros

# 

Para el estudio se hizo un análisis y limpieza de datos para la predicción de velocidad a partir de las humedades. Con esto, los resultados de la red LSTM en dos primeras instancias y en una tercera con GRU no se han conseguido en su cabalidad. Se ha visualizado que lo resultados obtenidos no son los deseados, por lo cual se debe seguir perfeccionando la red y analizando los datos.

8. Referencias

[1] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780. doi:10.1162/neco.1997.9.8.1735

[2] Zhao, Z., Chen, W., Wu, X., Chen, P. C., & Liu, J. (2017). LSTM network: a deep learning approach for short-term traffic forecast. IET Intelligent Transport Systems, 11(2), 68-75.

[3] Fu, R., Zhang, Z., & Li, L. (2016, November). Using LSTM and GRU neural network methods for traffic flow prediction. In 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC) (pp. 324-328). IEEE.

[4] S. Hochreiter, “Untersuchungen zu dynamischen neuronalen Netzen,” Dept. Fakultät für Informatik, M.S. thesis, Tech. Univ. Munich, Munich, Germany, 1991.

[5] S. Hochreiter, Y. Bengio, P. Frasconi, and J. Schmidhuber, “Gradient flow in recurrent nets: The difficulty of learning long-term dependencies,” in A Field Guide to Dynamical Recurrent Networks, S. C. Kremer and J. F. Kolen, Eds. Piscataway, NJ, USA: IEEE Press, 2001.

[6] A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, and J. Schmidhuber, “A novel connectionist system for unconstrained handwriting recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 5, pp. 855–868, May 2009.

[7] V. Pham, T. Bluche, C. Kermorvant, and J. Louradour. (Nov. 2013). “Dropout improves recurrent neural networks for handwriting recognition.” [Online]. Available: <http://arxiv.org/abs/1312.4569>

[8] P. Doetsch, M. Kozielski, and H. Ney, “Fast and robust training of recurrent neural networks for offline handwriting recognition,” in Proc. 14th Int. Conf. Frontiers Handwriting Recognit., 2014, pp. 279–284. [Online]. Available: http://people.sabanciuniv.edu/ berrin/cs581/Papers/icfhr2014/data/4334a279.pdf

[9] A. Graves. (Aug. 2013). “Generating sequences with recurrent neural networks.” [Online]. Available: <http://arxiv.org/abs/1308.0850>

[10] W. Zaremba, I. Sutskever, and O. Vinyals. (Sep. 2014). “Recurrent neural network regularization.” [Online]. Available: http://arxiv.org/ abs/1409.2329

[11] M.-T. Luong, I. Sutskever, Q. V. Le, O. Vinyals, and W. Zaremba. (2014). “Addressing the rare word problem in neural mac