

**PROPUESTA MODULO DE TALLER** (para aprobación por la Comisión de Carrera)

Nombre Actividad Específica	“Análisis de performance del método predictivo Graph WaveNet, en problemas de series de tiempo correlacionadas donde se busca la detección temprana de anomalías”
Proponente	Instituto de Computación, Facultad de Ingeniería, UDELAR
Responsable	Pablo Rodríguez Bocca ( <a href="mailto:prbocca@fing.edu.uy">prbocca@fing.edu.uy</a> )
Responsable en INCO o FING	Pablo Rodríguez Bocca ( <a href="mailto:prbocca@fing.edu.uy">prbocca@fing.edu.uy</a> )
Objetivo	<p>El <b>objetivo principal</b> es evaluar la performance (calidad y tiempo de ejecución) de la arquitectura Graph WaveNet de aprendizaje profundo en problemas de predicción en series de tiempo correlacionadas.</p> <p>Se alcanzará el objetivo principal mediante los siguientes siete objetivos específicos:</p> <ul style="list-style-type: none"> <li><b>O-1)</b> Familiarizarse con el problema de predicción de series de tiempo de NDVI (Normalized difference vegetation index), y su solución mediante arquitecturas convLSTM de aprendizaje profundo.</li> <li><b>O-2)</b> Desarrollo e implantación de una arquitectura Graph WaveNet en el CLUSTER.UY. Dicha implementación debe permitir la comparación justa con la técnica convLSTM.</li> <li><b>O-3)</b> Ejecución de pruebas en CLUSTER.UY. Las pruebas apuntan a comparar convLSTM con Graph WaveNet, respecto a calidad y tiempos de ejecución.</li> <li><b>O-4)</b> Documentar los resultados de los experimentos anteriores de performance.</li> <li><b>O-5)</b> Estudiar cómo utilizar las arquitecturas anteriores (convLSTM y Graph WaveNet) para la detección temprana de anomalías en la series de tiempo.</li> <li><b>O-6)</b> Ejecución de pruebas en CLUSTER.UY. Las pruebas apuntan a evaluar el desempeño de las arquitecturas para detectar anomalías.</li> <li><b>O-7)</b> Documentar los resultados de los experimentos anteriores de detección de anomalías.</li> </ul>
Descripción	<p>Existen varios satélites que proveen, en forma gratuita y periódica, imágenes de la superficie agrícola del país. A partir de estas imágenes se pueden obtener distintas medidas útiles para el sector productivo agrícola, como ser el NDVI (Normalized Difference Vegetation Index), que se relaciona con variables agronómicas de relevancia (índice de área foliar, déficit de nutrientes, afectación de pestes, etc) y con el rendimiento final del cultivo.</p> <p>Un modelo predictivo de NDVI puede ser muy útil para el sector productivo agrícola, puesto que permite detectar anomalías, es decir, desviaciones respecto del comportamiento esperado, las cuales pueden deberse a factores corregibles, como ser carencia de nutrientes/agua, o la proliferación de una peste.</p> <p>Desde un punto de vista formal, la predicción de NDVI se enmarca dentro del análisis predictivo de series de tiempo correlacionadas. Disciplina con un basto campo de aplicación, que es foco del proyecto I+D de CSIC titulado “Predicción de NDVI dentro de parcelas de cultivos agrícolas mediante técnicas de aprendizaje profundo”.</p>

	<p>En este proyecto trabajan tres investigadores de UDELAR, dos investigadores de E.E.U.U y un investigador de Francia. Este modelo taller se enmarca dentro de este proyecto. En el proyecto se han desarrollado varias técnicas predictivas, y se han evaluado en el problema particular de predicción de NDVI [1,2,3,4]. El objetivo de este módulo taller es comparar sus resultados con una nueva arquitectura de aprendizaje profundo, conocida como Graph WaveNet. Esta arquitectura ha sido evaluada en forma preliminar por el equipo del proyecto, obteniendo muy buenos resultados, los cuales pretendemos confirmar en este módulo taller.</p> <p>Por tanto, en el modulo taller, el estudiante trabajará en un problema bien definido, sobre los cuales se disponen de conjuntos de datos y varias metodologías validadas. En ese contexto, de trabajo en equipo, el estudiante del módulo evaluará exhaustivamente una nueva forma de solucionar el problema, la cual ha sido preliminarmente validada.</p> <p>Bibliografía:</p> <p>[1] Andrés Berger, Guillermo Ettlín, Christopher Quincke, Pablo Rodríguez-Bocca, “Predicting the Normalized Difference Vegetation Index (NDVI) by training a crop growth model with historical data”, Computers and Electronics in Agriculture, Available online 10 May 2018, ISSN 0168-1699, <a href="https://doi.org/10.1016/j.compag.2018.04.028">https://doi.org/10.1016/j.compag.2018.04.028</a>.</p> <p>[2] Rehaan Ahmad, Brian Yang, Guillermo Ettlín, Andrés Berger, cand Pablo Rodríguez-Bocca, “A Machine-Learning Based ConvLSTM Architecture For NDVI Forecasting”, International Transactions in Operational Research, To appear.</p> <p>[3] Assendorp, Jan Paul. 2017. "Deep Learning for Anomaly Detection in Multivariate Time Series Data." Master's thesis, HAW Hamburg, 2017.</p> <p>[4] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. Shroff. 2016. “LSTM-based encoder-decoder for multi-sensor anomaly detection”. In. 33rd International Conference on Machine Learning, Anomaly Detection Workshop. arXiv preprint arXiv:1607.00148.</p>
<p>Aporte a / tareas concretas del estudiante</p>	<p>El estudiante será el encargado de realizar la parte experimental, es decir las pruebas en el CLUSTER.UY, definidas en los objetivos <b>O-3</b> y <b>O-6</b>. Esto significa: instalar las herramientas necesarias, entender los experimentos a realizar, configurarlos en la infraestructura, ejecutarlos, controlar su correcta ejecución, y guardar los resultados.</p> <p>Para lograr lo anterior, el estudiante debe familiarizarse con el problema de estudio (objetivos específicos <b>O-1</b> y <b>O-5</b>), para lo cual tendrá el apoyo del equipo del proyecto, equipo que ha realizado varias publicaciones en la temática [1,2] y tiene previamente formas seleccionadas para evaluar la detección de anomalías [3,4]. El estudiante leerá la documentación disponible y participará de reuniones para entender la problemática.</p> <p>Dado que uno de los resultados de la experimentación es comparar tiempos de ejecución entre dos arquitecturas (convLSTM y Graph WaveNet). Esta comparación debe ser justa, asegurándose el acceso a los mismos recursos en el CLUSTER.UY (GPU, CPUs, memoria, etc.). Este es el objetivo <b>O-2</b>, donde el estudiante participará activamente, con el apoyo del equipo que tiene experiencia en esta infraestructura y el despliegue de la arquitectua convLSTM (a la cual agregaremos la arquitectura Graph WaveNet). El estudiante probará distintas configuraciones y realizará adaptaciones del código en búsqueda de una comparación lo más justa posible.</p> <p>La documentación de ambos experimentos (objetivos <b>O-4</b> y <b>O-7</b>) será realizado en</p>

	conjunto por todo el equipo del proyecto, dada la intención de transformar los mismos en publicaciones científicas. El estudiante participará en la documentación.
Carga horaria total	150 horas
Carga horaria sem.	20 horas / semana
Fecha inicio	8 de julio de 2020
Fecha fin	8 de septiembre de 2020
Conocimientos requeridos	Asignatura requerida: “Análisis de datos en redes complejas (grado)” o “Análisis de datos en redes (posgrado)”.
Cupo de estudiantes	1 o 2 estudiantes
Forma de Selección	Selección por escolaridad, requiriéndose conocimientos en análisis de datos en redes, modelado predictivo, y aprendizaje automático profundo. Se valorará la experiencia de uso del CLUSTER.UY y el conocimiento de la arquitectura Graph WaveNet
Método de Evaluación	Escrita. Se evaluará el documento final del trabajo, el cual se pretende presentar para su referato en una revista y/o conferencia científica relacionada.

\_\_\_\_\_  
Firma docente responsable  
inco – fing

aprobado Comisión Carrera fecha: