Escuela
Politécnica
Superior

Técnicas de aprendizaje automático aplicadas a la estimación del estado de cultivos mediante series temporales

Grado en Ingeniería en Sonido e Imagen en Telecomunicación

Trabajo Fin de Grado

Autor:
Anaida Fernández García
Tutor/es:
Juan Manuel López Sánchez
Tomás Martínez Marín

Abril 2020

Universitat d'Alacant
Universidad de Alicante

Técnicas de aprendizaje automático aplicadas a la estimación del estado de cultivos mediante series temporales

Autor
Anaida Fernández García

Tutor/es
Juan Manuel López Sánchez
Dpto. de Física, Ing. Sistemas y Teoría de la Señal
Tomás Martínez Marín
Dpto. de Física, Ing. Sistemas y Teoría de la Señal

Grado en Ingeniería en Sonido e Imagen en Telecomunicación

Escuela
Politécnica
Superior

Universitat d'Alacant
Universidad de Alicante

ALICANTE, Abril 2020

Justificación y Objetivos

Poner aquí un texto breve que debe incluir entre otras:

> "las razones que han llevado a la realización del estudio, el tema, la finalidad y el alcance y también los agradecimientos por las ayudas, por ejemplo apoyo económico (becas y subvenciones) y las consultas y discusiones con los tutores y compañeros. "

Agradecimientos

Es a ellos a quien dedico este trabajo.

dedicatoria

La distancia, que es el impedimento principal del progreso de la humanidad, será completamente superada, en palabra y acción. La humanidad estará unida, las guerras serán imposibles, y la paz reinará en todo el planeta.

Nikola Tesla.

Índice general

Índice de figuras

Índice de tablas

Índice de Códigos

Lista de Acrónimos y Abreviaturas

AEMA	Agencia Europea de Medio Ambiente.
IEEE	Institute of Electrical and Electronics Engineers.
NDVI	Índice de Vegetación de Diferencia Normalizada.
RAI	Real Academia de Ingeniería.
RF	Random Forest.
SAR	Synthetic Aperture Radar.
TFG	Trabajo Final de Grado.

1 Introducción

La telecomunicación se puede definir como toda transmisión y/o emisión y recepción de señales que representan signos, escritura, imágenes y sonidos o información de cualquier naturaleza por hilo, radioelectricidad, medios ópticos u otros sistemas electromagnéticos [1]. Esto permite compartir información útil a distancia y engloba un amplio conjunto de sistemas y tecnologías.

En este apartado nos vamos a centrar en situarnos dentro de los distintos sistemas de telecomunicación, y más detenidamente en los relevantes para este proyecto. A continuación, se expondrán los objetivos concretos que se quieren alcanzar. Y, por último, cómo se va a organizar la memoria del proyecto.

1.1 Contexto

Las telecomunicaciones forman parte de nuestro día a día y tienen cometidos de lo más variados: desde mandar un simple mensaje hasta comunicarse con una estación espacial, pero todos ellos engloban el manejo o el hecho de compartir información a distancia.

1.1.1 Tecnología

Dentro de los sistemas de telecomunicación encontramos la radio, la televisión, la telefonía fija y móvil, Internet por banda ancha o datos, la radionavegación o la teledetección. Todos ellos utilizan ondas electromagnéticas para sus comunicaciones, aunque estas se realicen mediante distintos medios de transmisión, que pueden ser guiados o no guiados, y con las modulaciones que se adapten a las necesidades de cada sistema.

Este proyecto se va a centrar en el sistema de la teledetección, definido como la adquisición de información un objeto, área o fenómeno, ya sea usando instrumentos de grabación o instrumentos de escaneo en tiempo real inalámbricos o que no están en contacto directo con el objeto, según la Real Academia de Ingeniería (RAI) [1]. Estos instrumentos van a medir la radiación electromagnética que emiten o reflejan los objetos observados. Algunos de estos instrumentos pueden ser cámaras fotográficas, láseres, sistemas de radar o sonar, y pueden ser pasivos, miden la radiación natural emitida o reflejada, o activos, emiten energía que posteriormente será reflejada y detectada.

Los instrumentos de medida, ya sean pasivos o activos, tienen la ventaja de poder estar situados a grandes distancias de la localización donde se quiera realizar la detección. Es por ello que se encuentran normalmente en satélites, aviones, barcos, etcétera, dependiendo de lo que se quiera medir. Las aplicaciones que engloba la teledetección son muy numerosas y suelen estar enfocadas a estudios científicos de ciertas áreas de la Tierra.

1.1.2 Caso particular a tratar

Una vez introducida la tecnología existente para el área de este proyecto, concretamos cuál va a ser nuestra situación.

Ya que la aplicación en la que se mueve este proyecto es la agrícola, concretamente la observación y adquisición de información de cultivos para su posterior estudio fenológico, la tecnología que se va a utilizar para ello son sistemas radar (radio detection and ranging), sistema activo, situado en un satélite artificial denominado Sentinel-1, del Programa Copérnico de la Agencia Europea de Medio Ambiente (AEMA). Estas tecnologías serán explicadas más detalladamente en el capítulo 2.

En este área ya hay estudios previos que, a partir de datos similares que comparten estos programas, se obtiene un estado de la fenología aproximado de los cultivos observados. Algunos estudios previos precedentes y que sirven de base para este Trabajo Final de Grado (TFG) son:

- [2], artículo de 2014 que trata de estimar el estado fenológico de cultivos en tiempo real empleando espacio de estados y técnicas de sistemas dinámicos utilizando información del pasado y actualizaciones y, finalmente una extensión del filtro de Kalman. La información que utiliza proviene de un radar polarimétrico del satélite Radsat-2 y los cultivos son 3 tipos de cereales.

- [3], artículo de 2016 que trata, de estimar el Índice de Vegetación de Diferencia Normalizada (NDVI), el cual representa el estado de la fenología, en tiempo real empleando filtros de partículas para integrar las dos fuentes de información utilizadas: imágenes Synthetic Aperture Radar (SAR) y temperatura del aire registrada. El satélite del que se obtiene la información es el TerraSAR-X y los cultivos observados son arrozales, como va a ser nuestro caso. Este obtienen resultados algo mejores que en el anterior artículo y se utiliza la misma tecnología que encontramos en este proyecto: SAR.

- [4], artículo de 2019 todavía más similar al objetivo de este proyecto, en él se estima el estado fenológico de distintos tipos de cultivos utilizando imágenes SAR proporcionadas por el satélite RADARSAT-2 y el método Random Forest (RF) para series temporales, que es uno de los elegidos también para este proyecto.

En resumen, para este proyecto en particular, el cultivo observado son arrozales, los datos empleados son imágenes SAR de los satélites Sentinel-1A y Sentinel-1B con ciclos periódicos de 6 días teniendo en cuenta ambos a partir de 2016, y las técnicas de estimación se basarán en las regresiones de series temporales y técnicas de aprendizaje automático.

1.2 Objetivos

Contribuyendo a la línea de investigación de los artículos [2] y [3], cuyos autores Juan Manuel López Sanchéz y Tomás Martínez Marín son el tutor y co-tutor de este TFG, respectivamente, el objetivo general sería estimar el estado de cultivos de arroz mediante el análisis series temporales con técnicas de aprendizaje automático y su unión a la línea de procesamiento original.

Los objetivos concretos serían:

- Analizar las posibles técnicas de regresión de aprendizaje autónomo (por ejemplo, regresión con RF) para estimar directamente el estado de los cultivos a partir de series temporales de datos.
- Analizar las posibles técnicas de aprendizaje autónomo para ser combinados con algoritmos ya disponibles de dinámica de sistemas en la estimación del estado de cultivos.
- Incorporar dichas técnicas en la cadena de procesado disponible.

1.3 Estructura de la memoria

La estructura de la memoria se va a dividir en 3 secciones principales las cuales son: marco teórico, metodología y resultados. Además de unas conclusiones finales valorando los resultados obtenidos.

En el marco teórico se expondrá toda la teoría necesaria para la compresión de este proyecto en términos técnicos y dentro de un contexto y una investigación previa que este continúa. Veremos en él las técnicas de regresión y machine learning existentes y candidatas para ser utilizadas, teoría de la teledetección, incluyendo cómo funcionan los sistemas radar, en concreto los SAR, qué información obtenemos y cómo interpretarla, y, finalmente, la estimación de parámetros físicos de los cultivos a partir de la información obtenida mediante regresión, qué parámetros son clave y qué procesamiento necesita la información para llegar a obtener estimaciones fiables y útiles.

En cuanto a la metodología, se incluirán tanto las técnicas y métodos concretos que se van a utilizar, por qué motivos y qué esperamos obtener de ellos, como el software, el lenguaje de programación que vamos a emplear, las herramientas utilizadas y las bases de datos con las que vamos a trabajar, incluyendo su procedencia y procesamiento previo.

Por último, el apartado de resultados expondrá los resultados obtenidos con las diferentes técnicas de regresión y aprendizaje automático para los mismos datos. Estos resultados podrán ser fácilmente evaluados ya que se contrastarán, además, con los datos reales tomados en tierra de los mismos cultivos que se presentan en el dataset.

2 Marco Teórico

A continuación se expone la teoría necesaria para la comprensión de este TFG, ampliando la información ya presentada en el capítulo 1.

2.1 Técnicas de regresión y machine learning

Las técnicas de regresión son todas aquellas técnicas que buscan la relación de una variable dependiente con una o más variables independientes mediante la estimación de su función de regresión. Para ello se consideran y ponderan todos los valores de la variable dependiente para unos valores fijos de las variables independientes. Además, en estos análisis, también se tiene en cuenta la varianza de la variable dependiente para estos mismos valores, pudiendo ser estudiada también mediante su distribución de probabilidad. Esta varianza indica la fiabilidad de nuestras estimaciones o el "ruido" en las medidas de la variable dependiente.

El caso más sencillo de regresión es en el que solo tenemos una variable dependiente y otra independiente, este caso se conoce como regresión lineal simple, ya que la función de regresión estimada se corresponde a una ecuación lineal de una recta. Los datos que obtenemos para la variable dependiente que vamos a relacionar tienen, aparte de las componentes lineales, una componente aleatoria de ruido que puede deberse a distintos fenómenos como la precisión mínima del instrumento de medida, el ruido que este mismo general en la medida o contribuciones de fuentes externas, consideradas como ruido también. Esta función de regresión es frecuentemente estimada mediante el método de mínimos cuadrados (MMC). También existe la regresión lineal múltiple, que funciona de la misma manera pero con mayor número de variables independientes, por lo que en lugar de una recta, la función de regresión representa un plano en el que coinciden N dimensiones, siendo N el número de variables independientes total.

Introducir aquí parte analítica regresión lineal

Cuando la función de regresión no es una función lineal, la regresión es no lineal, ya que la respuesta de la variable dependiente puede ser exponencial, logarítmica o polinomial, entre otras, por lo que la función de regresión presentará mayor complejidad. Aquí también es común utilizar el MMC o la regresión segmentada, que ajusta como regresión lineal segmentos de la original no lineal.

Cualquier variable independiente que tenga relación con la dependiente es útil en mayor o menor medida pero siempre proporciona información aunque su varianza sea muy grande o su contribución relativamente pequeña. Cualquier tipo de información extra proporciona un ajuste a la estimación final positivo si esta se ha modelado correctamente.

Introducir aquí parte analítica regresión no lineal

A parte de las regresiones lineales y no lineales mencionados, también encontramos otros métodos de regresión como son los mínimos errores absoluto (bastante similar al MMC), la regresión no paramétrica o la regresión lineal bayesiana.

Las técnicas de regresión proporcionan una estimación útil para realizar predicciones, por lo que están relacionados con el aprendizaje automático o machine learning. El machine learning es un tipo de inteligencia artificial, que se caracteriza por la generación de un modelo estimado de manera automática por un computador. Esta estimación se realiza con un entrenamiento previo aplicado a un algoritmo de aprendizaje específico a una serie de datos de entrenamiento. Con este aprendizaje se elabora un modelo que es capaz de devolver una salida o solución a partir de unos parámetros de entrada que deben ser del mismo tipo que los utilizados en la fase de aprendizaje.

* Introducir aquí esquema de ML *

Finalmente, el objetivo de las técnicas de machine learning puede ser clasificar una información o realizar una previsión acorde con un modelo estimado. Como se puede ver, el objetivo de este y las técnicas de regresión pueden coincidir y esto lleva a que tienen parte de su desarrollo en común.

2.1.1 Clasificación de técnicas de machine learning

Los modelos empleados en machine learning son numerosos, y su clasificación se puede realizar dependiendo de su algoritmo de aprendizaje y del tipo de razonamiento en el que se basa. Comenzando por la clasificación según su algoritmo de aprendizaje, que principalmente se dividen según el feedback del que aprenden, los modelos se pueden clasificar de la siguiente manera [5]:

- Aprendizaje no supervisado: este aprendizaje se basa en la clasificación o agrupación de los objetos de entrada según patrones que cumplen las distintas entradas de estos. Estos métodos no devuelven un nombre específico para cada grupo o cluster ya que no se le han proporcionado referencias o etiquetas en la etapa de entrenamiento. Necesita numerosas entradas en el entrenamiento para detectar patrones suficientemente estables.
- Aprendizaje por refuerzo: el aprendizaje se realiza por refuerzo positivo, que sería una recompensa, o negativo, penalización. Este algoritmo buscaría la estimación del modelo para obtener el máximo refuerzo positivo posible. Así, tras suficiente entrenamiento construye un modelo muy preciso para nuevas entradas.
- Aprendizaje supervisado: tanto las entradas como las salidas están previamente definidas en la etapa de aprendizaje. Se realiza un entrenamiento en el que se utilizan las entradas con sus correspondientes salidas para elaborar el modelo. Una vez suficientemente entrenado, este puede obtener salidas previamente desconocidas a partir de entradas similares a las del entrenamiento.
- Aprendizaje semi-supervisado: este aprendizaje recibe algunas de sus entradas correctamente etiquetadas y el resto de ellas, la mayoría, sin etiquetar, así tiene algunas

referencias para la clasificación fiables pero no toma las etiquetas como una referencia totalmente cierta para toda la clasificación como ocurre en el aprendizaje supervisado. Así se evitan malos aprendizajes por ruido o etiquetas erróneas en los datos de entrada. Bastante común en grandes masas de datos para aprendizaje.

Por otra parte, teniendo en cuenta la base de los razonamientos internos que los algoritmos realizan, aunque no considerando esta división estricta, las técnicas se pueden clasificar de la siguiente manera [6]:

- Geométricos: los modelos geométricos son aquellos cuyos objetos pueden ser representados en un espacio de instancias (X) en el que cada instancia corresponde a un posible objeto, esto es, habrá tantas instancias como objetos con distintas combinaciones de entradas posibles. Por otra parte, las etiquetas también se representan como un espacio de etiquetas (Y) con un número finito de posibilidades [7]. Utilizando estos conceptos, el algoritmo se desarrolla con otros conceptos geométricos como son líneas, planos y distancias. Estos métodos suelen ser aplicados cuando X e Y están formados por valores numéricos, que son fácilmente representables en ejes de coordinadas.
- Probabilísticos:
- Lógicos:

2.1.2 Modelos de machine learning y aplicaciones

2.2 Teledetección

2.3 Estimación de parámetros físicos de cultivos mediante regresión

Bibliografía

[1] de Ingeniería, R.A. *Diccionario español de ingeniería*, 2014. URL `http://diccionario.raing.es/es`.

[2] Vicente-Guijalba, F., Martinez-Marin, T., and Lopez-Sanchez, J.M. *Dynamical approach for real-time monitoring of agricultural crops*, 2014.

[3] Bernardis, C.D., Vicente-Guijalba, F., Martinez-Marin, T., and Lopez-Sanchez, J.M. *Contribution to real-time estimation of crop phenological states in a dynamical framework based on ndvi time series: Data fusion with sar and temperature*, 2016.

[4] Wanga, H., Magagia, R., Goïtaa, K., Trudela, M., McNairnb, H., and Powers, J. *Crop phenology retrieval via polarimetric sar decomposition and random forest algorithm. Elsevier*, 2019.

[5] Russell, S.J. and Norvig, P. *Artificial Intelligence A Modern Approach*, 2010.

[6] Flach, P. *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*, 2012.

[7] Fürnkranz, J. and Hüllermeier, E. *Preference Learning: An Introduction*, 2011.