



# Estimando el tamaño de polímeros con redes neuronales

Ekhine Irurozki

Basque Center For Applied Mathematics (BCAM)

Bilbao (Spain)

Email: eirurozki@bcamath.org

Jose A. Lozano

Intelligent Systems Group (ISG)

UPV/EHU, Donostia/San Sebastián (Spain)

Basque Center For Applied Mathematics (BCAM)

Bilbao (Spain)

Email: ja.lozano@ehu.eus

**Abstract**—En este artículo presentamos dos problemas de predicción asociados a procesos de fabricación avanzada en el contexto de la industria 4.0. Los dos problemas están fuertemente relacionados aunque pueden ser tratados de manera independiente.

El presente problema se centra en la fabricación de una pieza sólida de polímero a partir de una mezcla de materias primas en estado líquido. La anchura de la pieza resultante viene definida por el molde en el que las materias primas son vertidas. En cambio, la altura depende de la cantidad de materiales que son vertidos y la proporción de los mismos además de factores externos como la temperatura, etc. Desafortunadamente el proceso químico implicado es de tal complejidad que no puede modelarse de forma exacta y es difícil contestar a las cuestiones de interés, como (i) ¿cuál será la altura de la pieza dados estos parámetros de fabricación? y (ii) ¿cuáles son los parámetros de fabricación más adecuados para fabricar una pieza de una altura dada? Para ayudar en el proceso de fabricación y resolver estas cuestiones se ha llevado a cabo un estudio desde el ámbito de la ciencia de datos.

**Index Terms**—redes neuronales recurrentes, regresión, series temporales, fabricación avanzada, optimización

## I. INTRODUCTION

Nuestro problema se centra en una empresa del sector de la fabricación de polímeros sólidos. Las materias primas son una serie de líquidos y las piezas finales tienen forma de barra de sección rectangular. La generación de estos sólidos está regida por las leyes de la química pero las ecuaciones que describen el proceso no están definidas de manera exacta. Ligeros cambios en la combinación de las materias primas provocan grandes cambios en la forma final de la pieza. Además de estos parámetros, otros parámetros externos, como los factores ambientales de temperatura y humedad afectan al proceso de fabricación.

La empresa que nos ocupa, como muchas industrias modernas, está fuertemente sensorizada, por lo que recoge gran cantidad de datos del proceso de fabricación. Por ello y dados los problemas enumerados antes, se ha decidido afrontar el problema de la estimación de altura desde el enfoque de la ciencia de datos.

This work has been supported by the Spanish Ministry MINECO through BCAM Severo Ochoa excellence accreditation SEV-2013-0323 and the project TIN2017-82626-R funded by (AEI/FEDER, UE) and the ELKARTEK programs, and by the MINECO through BCAM Severo Ochoa excellence accreditation SVP-2014-068574.

Como en todo proceso industrial, la optimización de los recursos es crucial. En este caso concreto, es importante que la altura de las piezas fabricadas sea lo más cercana posible a la altura de las piezas encargadas por los clientes ya que el material en el excedente de altura tiene que ser descartado con el consiguiente coste asociado.

El proceso de fabricación tiene, a grandes rasgos, estas características. Existen diferentes tipos de formulaciones, cada una con una proporción de materias primas diferentes y que resultan en materiales de propiedades diferentes. Cambios en las proporciones de los componentes afectan a las diferentes propiedades, incluyendo la altura de la pieza. Como decíamos en párrafos anteriores, las condiciones ambientales juegan también un papel importante en las propiedades de las piezas resultantes.

Un experto decide las cantidades de los componentes y la máquina vierte la mezcla de componentes en una cinta transportadora en movimiento, de manera que la forma final de la pieza es de barra de sección rectangular. La anchura de la pieza viene determinada por unas paredes asociadas a la cinta transportadora, pero no así la altura.

Durante el proceso de fabricación de una pieza el operario puede modificar los parámetros de la máquina para que la pieza se ajuste a la altura deseada. Sin embargo, no es posible conocer con exactitud la altura de la pieza hasta que el proceso de fabricación ha terminado. Esto implica que el experto ha de estimar la altura de la pieza y hacer las correcciones en las cantidades de materias primas basándose en esa estimación.

La máquina de mezclado dispone de sensores que miden la cantidad de cada producto vertido, temperatura de un producto o la velocidad a la que se mezclan. Se realizan lecturas de estos sensores con una frecuencia determinada.

Al acabar de fabricar la pieza se pasa por unos sensores dimensionales que miden la altura cada  $M$  milímetros. Estos puntos donde se mide la altura se denominan *cortes*, concepto que se repite a lo largo del presente artículo.

Los cambios en las materias primas o en las condiciones ambientales afectan a la altura de la pieza de forma desigual y de manera difícil de determinar. Sin embargo, el enfoque de la ciencia de datos, que tan buenos resultados ha obtenido en diversos procesos industriales, ofrece una herramienta novedosa para afrontar estos retos.

Aunque la descripción de la base de datos con detalle se realiza más adelante es importante remarcar la representación de la misma como una colección de series temporales, cada una de ellas correspondiente a una pieza concreta. En otras palabras, cada pieza se describe por una serie temporal multivariada que se representa con una matriz bidimensional donde cada columna se corresponde con un parámetro de fabricación y cada fila con la lectura en un tiempo determinado. Los parámetros de fabricación son datos recogidos de sensores (como, por ejemplo, la cantidad de una materia prima concreta vertida en cada momento o la temperatura de la misma).

## II. FABRICACIÓN

En esta sección se detallan los problemas que se tratan de resolver y se describe la base de datos.

### A. Objetivos

Para mejorar la eficiencia de la fabricación de piezas hay dos puntos cruciales para reducir el gasto de material. Ambos son los objetivos a tratar en este artículo y se formalizan de la siguiente manera:

**O1:** Predecir la altura en cada uno de los cortes de la pieza dados los parámetros de fabricación.

**O2:** Predecir los parámetros de fabricación dada una altura de la pieza y unas condiciones ambientales de humedad y temperatura.

Los dos objetivos están relacionados estrechamente aunque cada uno por si solo resulta interesante.

### B. Data set, recogida y limpieza

Para crear la base de datos de las piezas es necesario juntar las variables de fabricación y de altura.

La base de datos está formada por los datos de 2504 piezas. Cada una de las piezas está descrita mediante una serie temporal multivariada con los valores recogidos por cada uno de los sensores de la máquina durante el tiempo en que dura el proceso de fabricación. La longitud de la pieza es variable y el número de cortes de las piezas de la base de datos está por lo general entre 100 y 200.

A lo largo de este artículo usaremos las denominaciones de serie temporal y secuencia de manera intercambiable.

El primer proceso a realizar en un contexto de ciencia de datos es el preproceso de la base de datos. En este caso el preproceso conlleva tanto tareas standard como otras diseñadas ad hoc para este problema concreto.

Los expertos aseguran que, con una probabilidad baja, los sensores de altura generan lecturas erróneas. Por ello, el preproceso comienza con la detección de errores en las lecturas de los sensores de altura según la definen los propios expertos. Una lectura en un corte  $p$ ,  $H_p$ , se considera errónea si  $|H_p - H_m| > 0.05H_{max}$ , donde  $H_m$  es la mediana de la altura en los 10 cortes anteriores y los 10 posteriores a  $p$  y  $H_{max}$ .

Una vez eliminadas las lecturas erróneas de sensores se eliminan aquellas variables que se mantienen constantes en todas las observaciones y para aquellas parejas de variables

que estén fuertemente correladas (más de 0.9) se elimina una de las dos variables.

Existen casos en los que una misma materia prima se distribuye en dos depósitos diferentes. Esta, puede salir con diferentes temperaturas por cada uno de los depósitos, o salir por uno de ellos y no por el otro. En este caso podemos unificar los datos de los diferentes tanques de la siguiente manera:

- La cantidad de producto total es la suma de las cantidades vertidas en los distintos depósitos.
- La temperatura total  $T$  viene dada por  $T = (m_1T_1 + m_2T_2)/(m_1 + m_2)$  siendo  $T_i$  la temperatura a la que se vierte desde el depósito  $i$  y  $m_i$  la cantidad que se vierte desde el depósito  $i$ .

Al final de este proceso hemos reducido considerablemente el número de variables a utilizar, desde las 215 iniciales a las 57 finales. Con el fin de proteger la confidencialidad, no se ofrecen detalles sobre las variables. En cualquier caso, se enumeran los distintos grupos de variables de que disponemos para poder entender mejor el resto del artículo.

- Variables **atmosféricas**, que incluyen humedad y temperatura.
- Variables de **fabricación** son todas aquellas que recogen los datos de los sensores de las materias primas, velocidad de la cinta transportadora, etc. Sobre estas variables el operario de la máquina puede actuar.
- Variables de **formulación**, variable categórica con 63 posibles valores diferentes, que define la composición base del material. Estos valores iniciales para los parámetros de fabricación son generalmente alterados durante la fabricación de la pieza.
- Variable de **altura** de la pieza.

## III. MODELADO MATEMÁTICO Y PREDICCIÓN

Como se describe en la Sección II-A, en este problema tenemos dos objetivos diferenciados. En los siguientes párrafos detallamos cada uno de los objetivos, los enfoques que utilizamos para resolverlos y las relaciones entre ambos.

### A. Objetivo O1

En el O1 se trata de predecir la secuencia de las alturas de la pieza en cada corte dada la serie con los parámetros de fabricación. Formalmente, hemos definido O1 como un problema de regresión de series temporales.

Una de las técnicas más habituales a la hora de tratar con series temporales es el de las *rolling windows* [2], que ya que permite el uso de técnicas clásicas de aprendizaje automático. Sin embargo, este enfoque no es válido para O1 por el hecho de que hasta que el proceso de fabricación de la pieza no ha sido completado no es posible medir la altura de la misma.

Otra alternativa muy popular en el ámbito del aprendizaje automático con series temporales es el de las redes neuronales, en particular en sus variantes de redes neuronales recurrentes (RNN por sus siglas en inglés).

Las RNN son una clase de redes neuronales capaces de hacer uso de información secuencial o temporal. Aunque las



RNN no se han usado habitualmente en la literatura de fabricación avanzada, son muy comunes en diversas áreas, muy especialmente en traducción automática [3], reconocimiento del habla [1] y reconocimiento de caracteres escritos a mano [8] entre otros.

Las redes neuronales básicas [5] logran mantener una *memoria* de los estados vistos con anterioridad mediante enlaces recurrentes en las neuronas internas. De esta manera, las salidas de un punto en la secuencia son tomadas como entrada en el siguiente punto de la secuencia y así se consigue capturar información de lo que ha sido calculado hasta ese momento.

Aunque en teoría la red puede almacenar información desde el principio de la secuencia, en la práctica solo es capaz de recordar unos pocos estados anteriores. Esto es debido a un fenómeno conocido como *vanishing/exploding gradients problem* por el que los pesos de la red no varían en el proceso de aprendizaje o tienden a infinito, ya que los gradientes que se calculan en los algoritmos de aprendizaje (backpropagation y métodos basados en gradientes) tienden a cero o tienden a infinito.

Las redes Long / short term memory (LSTM) se proponen [7] para intentar mitigar este problema y se han utilizado con éxito en diversos ámbitos [4]. Su diseño está inspirado en la circuitería más que en la biología ya que en cada una de las neuronas internas se definen puertas (una puerta de salida, otra de entrada y otra de borrar) y una célula de memoria. La función de las puertas es la de guardar la información de la célula de memoria permitiendo o bloqueando el flujo de información hacia ella.

En este trabajo consideramos ambos tipos de neuronas. Las arquitecturas consideradas se definen en la Sección IV.

### B. Objetivo O2

El segundo objetivo que se quiere cubrir en este artículo es el de estimar cuáles son los valores de los diferentes parámetros de fabricación para producir una pieza de un tipo (formulación) concreto, de una altura determinada y bajo unas condiciones atmosféricas concretas. Este objetivo lo enfocamos como un problema de optimización en el que tenemos dos componentes diferenciados, que son:

- el modelo de redes neuronales desarrollado en la sección anterior y
- una implementación del algoritmo *Covariance Matrix Adaptation Evolution Strategy* para la optimización numérica no lineal (CMA-ES por sus siglas en inglés).

Utilizando estos dos componentes, la estrategia que planteamos es utilizar el algoritmo de optimización CMA-ES para buscar en el espacio de parámetros de fabricación unos parámetros que, según el modelo de redes neuronales, produzcan una pieza con la altura deseada (o cuya altura se parezca lo más posible).

El algoritmo de optimización CMA-ES [6] es un método estocástico para la optimización numérica de funciones no lineares, multimodales, no convexas. Es un método evolutivo de caja negra y uno de los más utilizados en problemas de

alta dimensionalidad. El nombre viene del *covariance matrix adaptation* (CMA) un método para actualizar la matriz de covarianzas de la distribución.

Desde el punto de vista práctico, tres parámetros del algoritmo CMA-ES son necesarios para realizar una búsqueda con este algoritmo que son (i) la función objetivo a minimizar, (ii) una solución inicial y (iii) una varianza inicial en cada coordenada. Los valores específicos para estos parámetros se detallan en la Sección IV-B.

Tal y como hemos enunciado antes, los objetivos O1 y O2 están estrechamente relacionados dado el planteamiento que proponemos. Tanto que la solución para este segundo objetivo depende de los modelos propuestos para la solución del objetivo O1. De esta manera, la eficiencia en la resolución de este objetivo está condicionada por la calidad del algoritmo descrito en el objetivo O1.

## IV. EVALUACIÓN EXPERIMENTAL

En los siguientes apartados detallamos de qué manera se ha realizado la evaluación de las propuestas de este artículo y describimos los resultados que avalan esta propuesta.

### A. Resultados del objetivo O1

Para evaluar el modelo propuesto para la regresión en series temporales se han realizado una serie de experimentos con diversos parámetros

- *units* o número de neuronas, uno de [30, 50, 100, 150].
- *batch*, uno de [20, 50, 70, 90, 110].
- *model\_id*: cada uno de los modelos cuenta con una o dos capas con la siguiente arquitectura:
  - modelo 0: LSTM
  - modelo 1: LSTM con TimeDistributed layer
  - modelo 2: SimpleRNN
  - modelo 3: SimpleRNN con TimeDistributed layer
- *lenpart*: las piezas, que generalmente tienen entre 100 y 200 cortes, se particionan en subpiezas de *lenpart* cortes para que tengan todos la misma medida.
- *epochs*: 50
- *norm\_x*: valor 1 indica que los datos de las variables predictoras (las de fabricación y atmosféricas) se estandarizan y 0 que se normalizan.
- *norm\_y*: valor 1 indica que los datos de la secuencia a predecir (la de altura) se estandariza 0 que se normalizan.

La métrica utilizada en todos los casos es *mean squared error*, MSE. Se utiliza validación de 10 rodajas en cada una de las cuales se la base de datos se separa en los dos conjuntos, el grupo de entrenamiento y otro de test, de proporciones 0.7 y 0.3. Estos han de estar estratificados por tipo de formulación, de manera que las proporciones de cada tipo de formulación en las dos muestras sean constantes. Sin embargo, la variable de tipo de formulación no es relevante en ningún otro punto de la experimentación del objetivo O1, que sí utiliza las variables atmosféricas, de fabricación y altura.

En todos los experimentos realizados nos encontramos con que los primeros puntos son muy difíciles de aproximar. La red

necesita unos cuantos puntos para poder estimar con acierto la altura de la pieza. Por este motivo, a la hora de calcular el error en una red no se tiene en cuenta el error de los 10 primeros cortes.

Se aprende una red con cada una de las combinaciones de estos parámetros y se mide su error de entrenamiento y testeo durante las 50 épocas que dura el aprendizaje. El resultado después de 50 épocas se puede ver en la Figura 1, donde cada una de las 16 gráficas agrupa los experimentos con un mismo valor para los parámetros  $lenpart$ ,  $normx$  y  $normy$ .

Cada una de las ejecuciones se coloca en el punto X e Y se corresponden con los errores de entrenamiento y testeo. El color de cada punto indica el número de neuronas en el modelo.

Los ejes X e Y tienen los mismos valores en cada una de las gráficas para distinguir el sobreajuste de manera más sencilla.

Se puede ver que el comportamiento es similar para cualquier  $lenpart$  dados los mismos  $normx$  y  $normy$ . Se puede ver también que cuando se normalizan todos los datos la tendencia al sobreajuste es menor.

Tras un estudio general, observamos que los parámetros para los que estos modelos parecen comportarse mejor son  $lenpar = 50$ ,  $normx = 1$  y  $normy = 1$ . En la Figura 2 podemos ver los resultados de las ejecuciones para este conjunto de parámetros donde los colores de cada punto viene determinado por uno de los parámetros de la ejecución, de izquierda a derecha,  $batch$ ,  $units$  y  $model\_id$ . Se puede observar como los mejores resultados vienen dados por tamaños de  $batch$  pequeños, mayor número de neuronas y para los modelos 0 y 1.

Para analizar en detalle el comportamiento de la red proponemos la visualización de la Figura 3. Los parámetros de este experimento concreto son ( $units: 50$ ,  $batch: 20$ ,  $epochs: 100$ ,  $1\ model\_id$ ,  $60\ lenpart$ ,  $1\ normx$ ,  $1\ normy$ ). En el eje de las X están dispuesto los cortes de la pieza y en el de la Y la diferencia entre la altura real de la pieza y la altura estimada. Se han elegido 30 piezas de manera aleatoria del conjunto de entrenamiento y el error de cada uno de ellos viene dibujado con una línea de color diferente.

Se puede ver como los errores son casi constantes, no son picudos, y que la mayoría de ellos se agrupan entre los valores  $-0.02$  y el  $0.02$ .

## B. Resultados del objetivo O2

En esta sección describimos la configuración de la experimentación para este objetivo. Recordemos que encontrar un valor para cada una de las variables de fabricación de manera que junto con las variables ambientales generen una pieza de una altura dada. Para ello contamos con una red neuronal y con un algoritmo de optimización.

La red neuronal es la misma que hemos descrito en la sección anterior y que se visualiza en la Figura 3.

En el algoritmo de optimización hay que concretar una serie de valores, que son:

- La función objetivo a minimizar es la suma de la diferencia en cada corte entre la secuencia devuelta por la red y la altura real de la pieza.
- Una solución inicial, por cada variable de fabricación se toma la media de todos los cortes de todas las piezas del mismo tipo de formulación. Se consideran sólo aquellas piezas del conjunto de testeo.
- Una varianza inicial, por cada variable de fabricación se toma la varianza de todos los cortes de todas las piezas del mismo tipo de formulación. Se consideran sólo aquellas piezas del conjunto de testeo.

Para evaluar nuestra propuesta para este objetivo utilizamos el mismo conjunto de test que se utiliza para testar la red neuronal. Para una pieza dada tomamos sus variables de altura  $H$  y ambientales  $A$  de una pieza concreta. El reto es encontrar un conjunto de valores para cada una de las variables de fabricación  $F$  tales que, repetidas a lo largo del tiempo de fabricación, generen una pieza de una altura lo más parecida posible a la altura real de la pieza  $H$ . Para saber de qué altura sería la pieza fabricada con  $(F \cup A)$  utilizamos la red neuronal descrita anteriormente.

La Figura 4 muestra los resultados de nuestra propuesta. En la gráfica se muestra un grupo de barras por cada una de las variables de fabricación. Por motivos de confidencialidad los nombres de las variables (eje X) han sido eliminadas. Dentro de cada grupo de tres barras, de derecha a izquierda, cada barra representa (i) el valor medio de esa variable en todos los cortes de la pieza, (ii) el valor óptimo que el algoritmo ha encontrado y (iii) el valor de la solución inicial para esa variable.

Como se puede ver, las soluciones iniciales pueden parecerse más a la media de la variable utilizada. Sin embargo, según el modelos de redes neuronales el error en la altura de la pieza generada es muy pequeño. No ha sido posible todavía probar los valores optimizados en el mundo real.

La Figura ?? muestra en detalle ese mismo experimento. En ella se muestra la altura real de la pieza, la altura predicha por el modelo de redes neuronales para los parámetros de fabricación y la altura predicha por el modelo de redes neuronales para los parámetros óptimos encontrados.

Esta búsqueda ayuda a unificar los criterios de los diferentes expertos y arroja luz sobre la manera en la que las diferentes alteraciones de componentes o combinaciones de componentes afectan a la altura de la pieza generada.

## V. CONCLUSIONES

En este artículo presentamos dos novedades en el ámbito de la industria 4.0.

La primera es un modelo para predecir la altura de una pieza fabricada basándose en sus variables de fabricación (materias primas) y en las condiciones ambientales. Este modelo está construido con redes neuronales, en particular con redes neuronales recurrentes debido a su capacidad de trabajar con series temporales.

La segunda es un algoritmo de optimización que busca los parámetros de fabricación óptimos para fabricar piezas



Fig. 1. Resultados de la experimentación intensiva con las distintas redes neuronales.

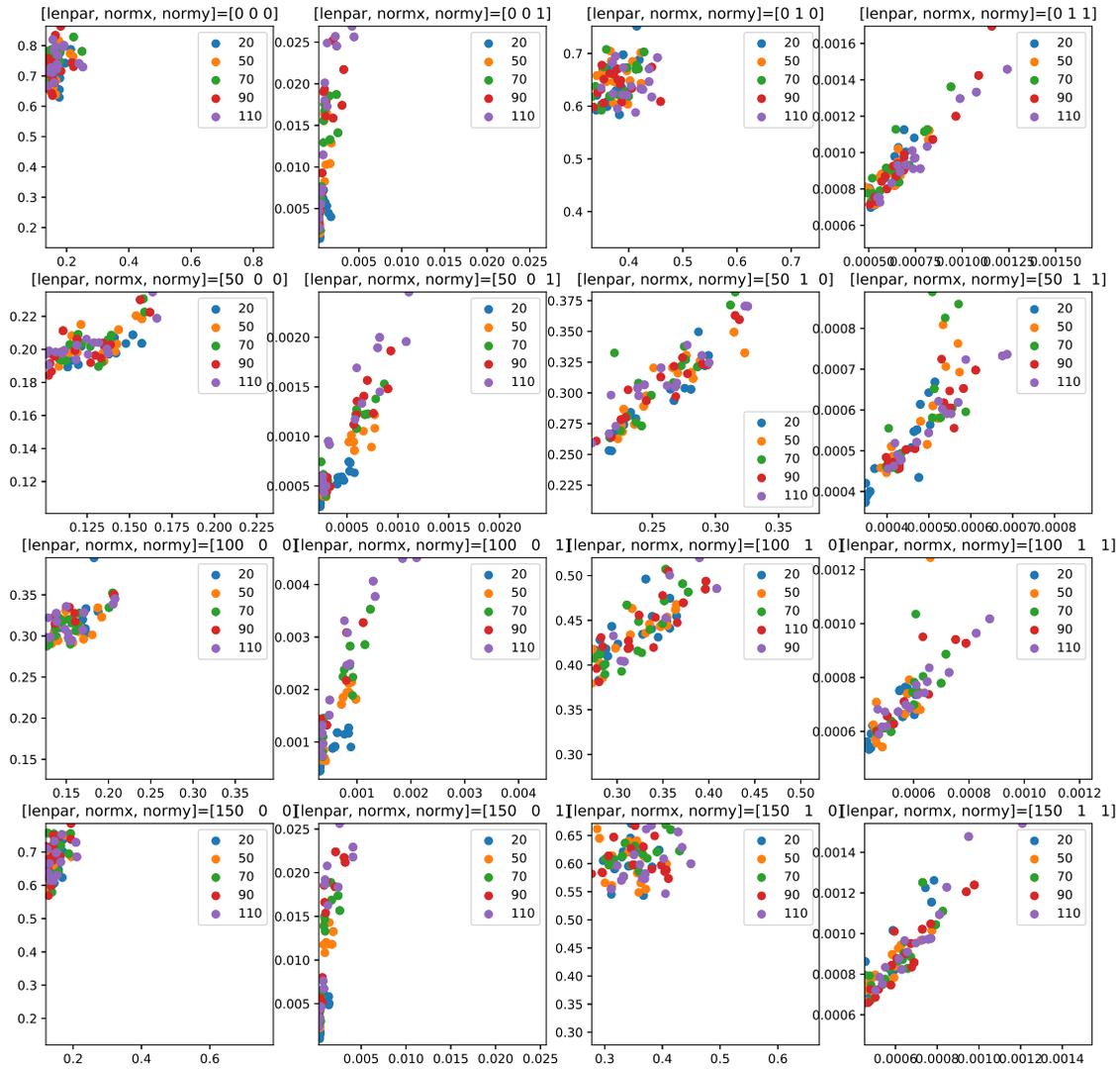


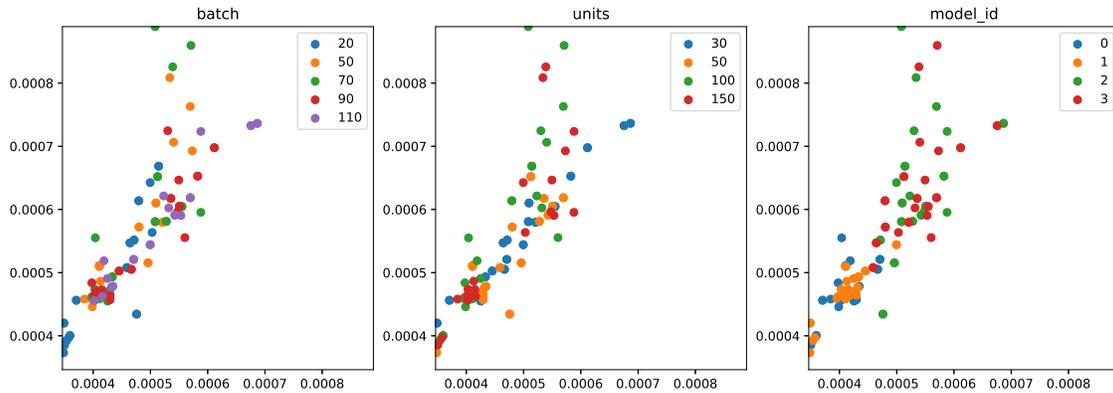
Fig. 2. Detalle de los experimentos con  $lenpar = 50$ ,  $normx = 1$  y  $normy = 1$ .

Fig. 3. Diferencia entre la altura real de la pieza y la predicha por nuestro algoritmo.

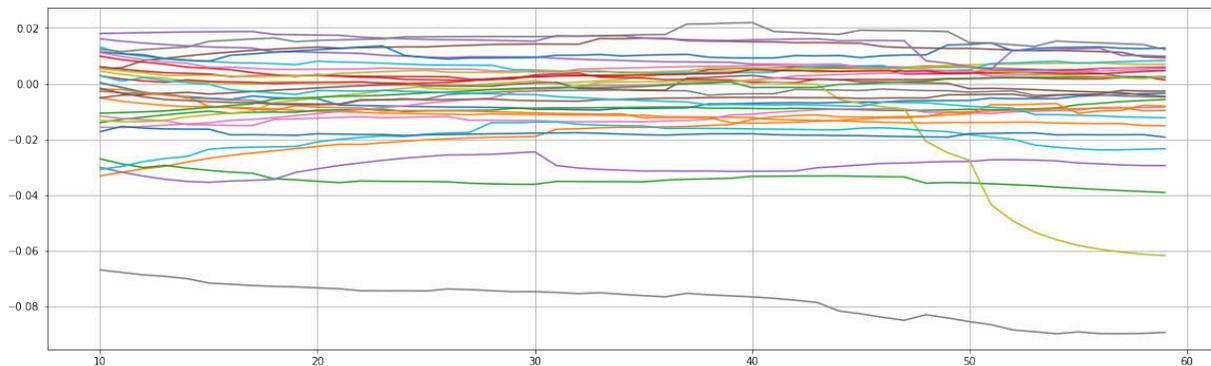
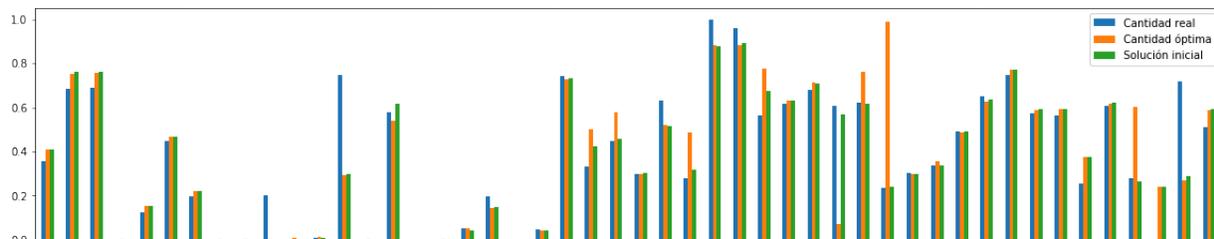


Fig. 4. Cantidad de materias primas reales y estimadas, tanto como solución inicial como optimizadas.



de una altura determinada bajo unas condiciones ambientales concretas. Esta búsqueda se efectúa utilizando el modelo de redes neuronales descrito anteriormente en este mismo artículo.

Las pruebas experimentales demuestran la eficacia de este enfoque y abre la puerta a futuros trabajos del uso de RNN en la industria para regresión en la industria de la fabricación avanzada.

#### REFERENCES

- [1] A.V. Allin Geo and K.P. Kaliyamurthie. Discourse recognition with deep learning. *International Journal of Pure and Applied Mathematics*, 116(10 Special Issue), 2017.
- [2] Peter Brockwell and Richard Davis. *Introduction to Time Series and Forecasting*. 2002.
- [3] Aaron Goodfellow, Ian, Bengio, Yoshua, Courville. Deep Learning (Chinese Translation). *MIT Press*, 2016.
- [4] Xiangang Li and Xihong Wu. Constructing Long Short-Term Memory based Deep Recurrent Neural Networks for Large Vocabulary Speech Recognition. *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4520–4524, 2014.
- [5] L R Medsker and L C Jain. RECURRENT NEURAL Design and Applications. *New York*, 2001.
- [6] Nikohansen. pycma. [\url{https://github.com/CMA-ES/pycma}](https://github.com/CMA-ES/pycma), 2017.
- [7] Long Short and Term Memory. Long Short Term Memory. *Memory*, (1993):1–28, 1996.
- [8] Sheng Zheng, Xiangyun Zeng, Ganghua Lin, Cui Zhao, Yongli Feng, Jinping Tao, Daoyuan Zhu, and Li Xiong. Sunspot drawings handwritten character recognition method based on deep learning. *New Astronomy*, 45:54–59, 2016.