

Tarea Final



José Clavijo

Ernesto Bazzano

**Curso: Introducción a la Ciencia de
Datos Facultad de Ingeniería – Udelar.**

Contenido

1. Juego de Datos	2
1.1 Que características tiene el juego de datos?	2
1.2 Que problemas nos podemos encontrar en el juego de datos?	2
2. Que problemas/preguntas pueden ser resueltos /contestados utilizando las herramientas presentadas en el curso?	3
3. Como pueden ser resueltos/contestadas los/las Problemas/Preguntas?	3
3.1 Plan de Producción.....	3
3.2 Entrenamiento de Operarios.....	5

1. Juego de Datos

De una empresa industrial, se cuenta con el Plan de Ventas y con el Plan de Producción (secuencia ordenada de órdenes de producción) desde 2013 a 2023.

Dentro del plan de ventas se cuenta con la siguiente información:

- Características del producto vendido.
- Cantidad de unidades vendidas.
- Fecha de Entrega.

Dentro del plan de producción ejecutado se cuenta con la siguiente información por orden de producción:

- Máquina dónde fue ejecutada la orden de producción.
- Operario que realizó la producción.
- Horario de inicio, horario de fin de producción: tiempo total de producción.
- Velocidad de producción.
- Tiempos de paradas previstas e imprevistas.
- Cantidad de Unidades de producto producidas.
- Cantidad de Unidades descartadas (para calcular desperdicio).
- Rendimiento objetivo para la orden (tiempo, desperdicio).
- Velocidad objetivo de la máquina.

1.1 Que características tiene el juego de datos?

- La información está disponible en una planilla Excel, en formato de tabla, la cual puede ser exportada para tratar con librerías pandas o numpy.
- La mayoría de los datos son numéricos, excepto el operario de la máquina(es un texto con el nombre del operario), el cual debe ser transformado a un número contemplando su nivel de desarrollo y desempeño, y las fechas, las cuales en principio pueden ser transformadas fácilmente a formato numérico.

1.2 Que problemas nos podemos encontrar en el juego de datos?

Los datos son generados desde el puesto de trabajo operativo, podemos encontrarnos con los siguientes problemas:

- **Datos Faltantes Parcialmente:** No aparecen determinados tiempos y/o datos del proceso, ejemplos: no aparece el tiempo de descanso operativo o la velocidad de producción.
- **Datos inválidos:** tiempos negativos, tiempos fuera de lo esperable.
- **Falta de Continuidad del plan de producción:** la fecha/hora de fin de producción de una orden debe coincidir con la fecha/hora de inicio de producción de la siguiente orden, en ocasiones no coinciden.
- **Datos Faltantes Totales:** no disponer de toda la información asociada a una orden de producción.
- **Tiempos de paradas de máquina no explicado:** este es un problema importante, porque estos tiempos pueden estar relacionados a paradas imprevistas y en consecuencia se

pierde información valiosa para el análisis. Habría que ver si es posible reconstruir esta pérdida de información con alguna regla o criterio.

2. Que problemas/preguntas pueden ser resueltos /contestados utilizando las herramientas presentadas en el curso?

1. Dado un plan de ventas, cual es el plan de producción óptimo para cumplir con la entrega de los productos vendidos?
2. Dado un operario en etapa de entrenamiento, cuál será su desempeño esperado al finalizar el entrenamiento?

En la próxima sección se desarrollará una repuesta a estas preguntas utilizando, dentro de las herramientas brindadas en el curso, algoritmos de aprendizaje automático.

No se desarrolló una respuesta a las siguientes preguntas, entendemos que también podrían contestarse utilizando herramientas del curso:

3. Dado un nuevo producto a producir, cuál será el mínimo y máximo costo de producción(entendiendo como costo, el tiempo de producción)?
4. Los clientes envían un pronóstico de compra de productos para X períodos, cuál será el pronóstico ampliado para los períodos $X+1, X+2, \dots, X+N$?
5. Dada la última falla imprevista por rotura de una máquina, cuándo se espera que ocurra la próxima falla?.

3. Como pueden ser resueltos/contestadas los/las Problemas/Preguntas?

3.1 Plan de Producción.

Para responder esta pregunta, es necesario definir el problema a resolver:

Dado un plan de ventas, el cual generará **N órdenes de producción**, la tarea a resolver implica asignar estas órdenes a un conjunto de **máquinas** (de **M** máquinas disponibles), con el objetivo de **reducir el tiempo** total de **producción**, sujeto a que se dispone de una **capacidad máxima de producción** para cada máquina. Esta asignación óptima de órdenes a máquinas resulta en el plan de producción óptimo.

Para evaluar dónde los algoritmos de Aprendizaje Automático pueden ser de utilidad para la resolución de este problema, vamos a tener que conocer cuál es la secuencia de tareas a realizar para armar un plan de producción óptimo y evaluar que tareas se podrían realizar con algoritmos de Aprendizaje Automático. Antes de avanzar en esto, a continuación vamos a comenzar describiendo los atributos de un producto y cuáles son las entradas/salidas del sistema.

- **Atributos de un Producto:** cada producto tiene 3 atributos principales a los que llamaremos {D,G,e}.
- **Clase de Producto:** dos productos pertenecen a la misma clase si los tres atributos principales son iguales.
- **Entrada al sistema:** Secuencia de N órdenes de producción, cada orden se puede representar como un vector de 2 componentes (clase de producto(c), cantidad a producir(p)).
- **Salida del sistema:** A cada orden de producción se le asigna una Máquina (m) y un tiempo de producción (t_p).

A continuación se puede observar la secuencia de tareas a realizar,

- **Secuencia de Tareas – Resolución de Problema Combinatorio.**
 - 1) **Agrupar y Ordenar órdenes de producción:** agrupar las órdenes de producción que tengan igual atributo **D**, posteriormente ordenar por atributo **e** y finalmente ordenar por atributo **G**. Con esta tarea obtenemos conjuntos de órdenes agrupadas y ordenadas $D_c = [D_1, D_2, \dots, D_{nD}]$, n_D : es la cantidad de productos con atributo D diferente.
Esta tarea tiene como objetivo disminuir el tiempo de producción al seguir estas reglas de agrupamiento y ordenamiento para esta clase de productos a producir.
 - 2) **Espacio de Soluciones:** Calcular t_p para todas las combinaciones posibles de asignación $(D_1, D_2, \dots, D_{nD}) \leftrightarrow m$
Ejemplo: si tengo dos elementos en el conjunto $D_c = \{D_1, D_2\}$ y dos máquinas $M = (m_1, m_2)$. El espacio de soluciones sería el siguiente:
 - a) $(D_1, D_2) \leftrightarrow m_1$: conjunto de órdenes D_1 y D_2 se asignan a la Máquina 1.
 - b) $(D_1, D_2) \leftrightarrow m_2$
 - c) $D_1 \leftrightarrow m_1$ & $D_2 \leftrightarrow m_2$
 - d) $D_2 \leftrightarrow m_1$ & $D_1 \leftrightarrow m_2$
 Para el cálculo son utilizados tiempos de producción esperados fijos, calculados con información relacionada a la velocidad de la máquina y eficiencia esperable.
 - 3) **Espacio de Soluciones Factibles / Solución Óptima:** Para alcanzar el espacio de soluciones factibles, hay que eliminar todas las soluciones que no cumplen la restricción del problema: el tiempo acumulado de producción de una máquina supera la capacidad de producción establecida. Entonces, del espacio de soluciones factibles obtenido, seleccionamos la solución con menor tiempo de producción.

Uso de Algoritmos de Aprendizaje Automático:

Cuando el espacio de soluciones es muy grande, las tareas 2) y 3) no son tareas eficientes del punto de vista computacional, se pueden usar otras reglas de asignación de órdenes que no aseguran lograr el óptimo global pero si aseguran disminuir los tiempos de encontrar una solución factible, por ejemplo el uso de Métodos Heurístico, Algoritmos Evolutivos, o Algoritmos de Aprendizaje Automático.

Como nuestros datos de entrada al sistema son una secuencia de órdenes de producción, se podría utilizar una red neuronal recurrente (RNN) para resolver el problema, ya que estas redes tienen la particularidad de trabajar con secuencia de datos. La red debe ser entrenada con el siguiente conjunto de datos:

- X (entrada) = Secuencia de órdenes de producción agrupadas y ordenadas $D_c = [D_1, D_2, \dots, D_{n_p}] \rightarrow$ si D_j puede ser representado por n_f features y n_p es la cantidad de conjuntos a evaluar (time steps), podemos representar la entrada como un tensor $(1, n_p, n_f)$ (Representación de la entrada de la red RNN).
- Y (salida) = Secuencia de asignación “máquina(m)/tiempo de producción(t_p)” para el conjunto D_c . $(m, t_p) = [(m, t_p)_1, (m, t_p)_2, \dots, (m, t_p)_{n_p}] \rightarrow$ esta secuencia puede representarse como un tensor $(1, n_p, 2)$ (Representación de la salida de la red RNN).

Posterior al entrenamiento, dado un nuevo conjunto de órdenes de producción agrupadas y ordenadas $D_c^{\text{NUEVO}} = [D_1, D_2, \dots, D_{n_p}]$, mediante la red RNN se puede encontrar una asignación: $[(m, t_p)_1, (m, t_p)_2, \dots, (m, t_p)_{n_p}]$ a este conjunto de órdenes. Posteriormente, agrupando para cada una de las máquinas, los conjuntos de órdenes D_j asignados a ellas, se puede obtener el plan de producción con el tiempo de producción total por máquina estimado.

Hay un punto del problema que no fue incorporado explícitamente en la solución planteada con la red neuronal RNN, el cumplimiento de la restricción de capacidad. Se debe estudiar como incorporar la restricción al algoritmo para garantizar la factibilidad de la solución brindada.

3.2 Entrenamiento de Operarios

En el dataset se pueden identificar órdenes de producción de distintos operarios ya entrenados durante su etapa de entrenamiento. Para este conjunto de órdenes se podrían extraer dos datos: 1) tiempo real de producción de la orden y 2) tiempo esperado de producción de la orden. Con estos datos se puede calcular una medida de eficiencia como la relación entre ellos. Entonces, se pueden agrupar órdenes realizadas durante el entrenamiento por un mismo operario para un período de tiempo fijo (por ejemplo: semanal, mensual) y calcular este valor de eficiencia para cada período del entrenamiento, obteniéndose una secuencia de valores de eficiencia, que sería una medida de la evolución del desempeño del operario durante el entrenamiento.

Si P es la cantidad de períodos establecidos para finalizar el entrenamiento y $p < P$, es la cantidad parcial de períodos de entrenamiento alcanzados por un operario, se necesita entrenar un algoritmo que, dada una secuencia de eficiencias obtenidas en los primeros p períodos, pueda predecir el nivel de eficiencia que tendrá el operario en los próximos $(P-p)$ períodos de entrenamiento.

Se podría resolver como un problema de predicción de series temporales en tiempo discreto, utilizando una red neuronal recurrente (RNN, LSTM) que tiene como entrada una secuencia de datos, en este caso la secuencia sería los valores de eficiencia obtenidos durante los primeros p períodos del entrenamiento, y que nos devuelva $(P-p)$ valores de eficiencia esperados a futuro.

El fin sería detectar, en forma temprana, entrenamientos que requieran tomar acciones para mejorar la evolución de estos.