

Predicción de Inductancia en datos de un Inversor Fuente de Voltaje (VSI)

Sergio A. Silva-Rubio

1. Introducción

Este informe busca replicar los estudios realizados en [1, 2], los cuales estudian la predicción de inductancia en datos de un Inversor Fuente de Voltaje (VSI). Esta potencia es utilizada por un motor que opera a diferentes velocidades, generando cargas variables. El objetivo es asegurar el correcto funcionamiento de un VSI utilizando la estrategias de control llamada control predictivo. Para la predicción de inductancia en datos de un VSI se probaron varios algoritmos de clasificación, los cuales fueron evaluados y comparados.

2. Metodología

El conjuntos de datos consta de 399 instancias (filas), los cuales fueron divididos en dos conjuntos de datos 359 para datos de entrenamiento y 40 para datos de prueba.

2.1. Transformación de datos

Se realizó un cambio sobre los nombres de las variables tanto en los datos de entrenamiento como en los de prueba. Los cuales fueron enumerados desde 1 a 5000.

```
names(data.train) <- c(1:5000)
names(data.test) <- c(1:5000)
```

2.2. Modelos predictivos de clasificación

2.2.1. K-Nearest Neighbor (KNN)

Uno de los métodos de clasificación más fundamentales y simples. La clasificación K-vecino más cercano se desarrolló a partir de la necesidad de

realizar un análisis discriminante cuando las estimaciones paramétricas confiables de las densidades de probabilidad son desconocidas o difíciles de determinar. Ha utilizado a menudo en estos problemas de reconocimiento de patrones.

2.2.2. Support Vector Machine (SVM)

Es una herramienta de predicción de clasificación y regresión que utiliza la teoría de aprendizaje de la máquina para maximizar la precisión de la predicción mientras que evita automáticamente el sobreajuste de los datos. Cuando las clases no son perfectamente separables podemos aumentar la dimensionalidad de una manera específica, mediante el uso de kernels. Los kernels son funciones que transforman un espacio de pocas dimensiones en un espacio de dimensiones mayores mediante transformaciones complejas de los datos. En específico en este informe se probaron dos:

SVM Radial: El kernel radial tiene un comportamiento muy local, en el sentido de que solo las observaciones de entrenamiento cercanas a una observación de test tendrán efecto sobre su clasificación.

SVM Lineal: El kernel lineal cuantifica la similitud de un par de observaciones usando la correlación de Pearson. Con un kernel lineal, el clasificador obtenido es equivalente a un support vector classifier.

2.2.3. Naïve Bayes (NB)

Este algoritmo llamado *naïve* (ingenuo en español), simplifica enormemente el aprendizaje asumiendo que las características son independientes entre sí. Aunque la independencia es generalmente una suposición, en la práctica logra competir bien con clasificadores más sofisticados.

2.2.4. Decision Tree (DT)

Un árbol de decisión es un modelo de predicción que dado un conjunto de datos se fabrican diagramas de construcciones lógicas, que sirven para representar y categorizar una serie de condiciones que ocurren de forma sucesiva, para la resolución de un problema.

2.2.5. Random Forest (RF)

Es una combinación de árboles de decisión tal que cada árbol depende de los valores de un vector aleatorio probado independientemente y con la misma distribución para cada uno de estos.

2.3. Arquitectura

1. La primera etapa consiste cargar los datos y cambiar el nombre de las variables.
2. La segunda etapa corresponde al entrenamiento de los datos, que nos permite encontrar relaciones, desarrollar comprensión, tomar decisiones y evaluar su confianza a partir de los datos de entrenamiento. Además, para este caso contamos con información que nos permite clasificar los datos mediante una etiqueta.
3. La tercera etapa consiste en seleccionar los algoritmos que usaremos para predecir los valores de salida y así realizar una comparación entre ellos.
4. A partir de la tercera etapa podemos crear nuestro modelo predictivo que a con nueva información podrá clasificar en alguna de las etiquetas previamente definidas.

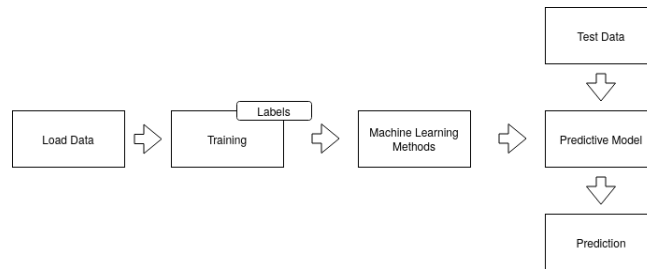


Figura 1: Arquitectura de experimentos.

2.4. Evaluación y Comparación de los modelos

Para problemas de clasificación estadística, una matriz de confusión es una herramienta que permite la visualización del desempeño de un algoritmo que se emplea en aprendizaje supervisado. Cada columna de la matriz representa el número de predicciones de cada clase, mientras que cada fila representa a las instancias en la clase real.

3. Resultados

Los resultados se obtuvieron tras ejecutar cada método utilizando los mismos datos de entrenamiento y prueba. Estos datos fueron cargados solo

una vez.

Method	Accuracy
KNN	0.8461538
SVM-R	0.9230769
SVM-L	0.8717949
NB	0.8974359
DT	1
RF	0.9487179

4. Conclusión

Se logró implementar seis métodos los cuales fueron comparados utilizando los mismos parámetros. Esto permitió determinar el desempeño de los algoritmos analizados a través de una comparación justa. En relación a los resultados se obtuvo un desempeño perfecto en el árbol de decisión, por contraparte KNN obtuvo un peor resultado. Esto se puede deber a la forma de seleccionar los datos debido a que solo fueron cargados una vez. Una mejora para esto, sería realizar varias pruebas utilizando distintos conjuntos de entradas, y a partir de ello se podría calcular un promedio.

Referencias

- [1] Aldana, D., Salgueiro, Y., Bellinger, C., Rivera, M., .& Astudillo, C. A. *Data for resistance and inductance estimation within a voltage source inverter*. Data in brief, 25. 2019.
- [2] Aldana, D., Salgueiro, Y., Bellinger, C., Rivera, M., & Astudillo, C. A. *Performance assessment of classification methods for the inductance within a VSI*. IEEE International Conference on Automation/XXIII Congress of the Chilean Association of Automatic Control (ICA-ACCA). Concepcion, Chile, 2018, p. 2018.