

Recomendación de Modelos Predictivos para Diagnóstico de Vibraciones en Turbogeneradores

Resumen

Se requiere desarrollar un sistema de clasificación automática del estado de turbogeneradores mediante análisis de vibraciones, con capacidad de diagnóstico de fallos. Se recomienda una estrategia gradual desde modelos interpretables hasta arquitecturas avanzadas. La lista priorizada es: Árboles de Decisión/Random Forest, Gradient Boosting Machines (XGBoost/LightGBM), Redes Neuronales Artificiales (PSO-BP), y SVM con kernel RBF. La recomendación principal es XGBoost por equilibrio entre rendimiento, interpretabilidad y eficiencia computacional, respaldado por evidencia en aplicaciones similares (Zhang & Su, 2010; Huang et al., 2004).

Selección de Modelos

Árbol de Decisión y Random Forest

Los árboles de decisión operan como un sistema experto que hace preguntas secuenciales sobre los datos de vibración. Comienza examinando la variable más importante, típicamente el valor RMS de velocidad, y pregunta: "¿Es mayor a 4.5 mm/s?" Según la respuesta, sigue por diferentes ramas haciendo preguntas más específicas como: "¿La relación entre la amplitud a 2X y 1X es mayor a 0.3?." "¿La componente axial es significativa?"

Cada camino a través del árbol representa una combinación específica de condiciones que lleva a un diagnóstico particular. Por ejemplo, si encuentra que el RMS es alto, la relación 2X/1X es elevada y la componente axial es dominante, concluye "desalineamiento probable". La belleza de este método es su transparencia - puedes seguir exactamente el razonamiento que llevó al diagnóstico.

Los bosques aleatorios mejoran esto creando cientos de árboles, cada uno especializado en diferentes aspectos de los datos. Un árbol puede enfocarse en patrones espectrales, otro en relaciones entre armónicos, y otro en la influencia de parámetros operativos. Al combinar sus votos, el sistema se vuelve más robusto y capaz de manejar casos complejos donde múltiples fallos pueden presentarse simultáneamente.

Ventajas:

- Manejan automáticamente valores faltantes mediante división óptima
- No requieren normalización de características
- Proporcionan importancia de características interpretable

- Robustos a valores atípicos y relaciones no lineales

Riesgos:

- Sobreajuste si la profundidad es excesiva
- Sensibles a desbalanceo de clases
- Pueden perder precisión con relaciones lineales simples

Preparación de datos:

- Codificación label encoding para variables categóricas
- No se requiere escalado de características
- Imputación simple para valores faltantes (mediana/moda)
- Feature engineering: ratios espectrales (2X/1X), estadísticos temporales

Gradient Boosting Machines (XGBoost/LightGBM)

XGBoost funciona como un equipo de especialistas que se entrenan secuencialmente, donde cada nuevo experto aprende de los errores del anterior. El primer "experto" hace una diagnosis inicial basada en las características más obvias, como el valor RMS. El segundo experto examina los casos donde el primero se equivocó, quizás notando que ciertos patrones espectrales fueron ignorados.

Este proceso continúa a través de cientos de ciclos, con cada nuevo modelo enfocándose en los casos más difíciles de diagnosticar. Por ejemplo, puede aprender que cuando la carga operativa está por encima del 80 por ciento, los umbrales de vibración deben ajustarse, o que ciertas combinaciones de armónicos indican fallos incipientes en rodamientos antes de que sean detectables por métodos tradicionales.

Lo que hace a XGBoost particularmente efectivo para vibraciones es su capacidad de capturar interacciones complejas entre variables. Aprende no solo que una alta amplitud en 2X sugiere desalineamiento, sino que cuando esto ocurre junto con un aumento específico en la temperatura del aceite y ciertos patrones de fase, la confianza en el diagnóstico aumenta significativamente.

Ventajas:

- Alta precisión en problemas tabulares complejos
- Manejo nativo de valores faltantes
- Regularización incorporada contra sobreajuste
- Eficiente con grandes volúmenes de datos

Riesgos:

- Mayor costo computacional en entrenamiento
- Sensibilidad a hiperparámetros

- Requiere tuning cuidadoso para óptimo rendimiento

Preparación de datos:

- Codificación label encoding para categóricas
- Normalización opcional (mejora convergencia)
- Feature engineering: componentes espectrales, estadísticos de rodamiento
- Validación temporal para evitar data leakage

Redes Neuronales Artificiales (PSO-BP)

Las redes neuronales imitan el procesamiento de información del cerebro humano para analizar patrones complejos en los datos de vibración. La versión PSO-BP combina dos fases: primero usa optimización por enjambre para encontrar una buena configuración inicial, y luego refina los parámetros mediante aprendizaje supervisado.

En la práctica, la red neuronal toma todas las características de vibración - valores RMS, espectros de frecuencia, parámetros operativos - y las procesa a través de múltiples capas de "neuronas artificiales". Cada capa detecta patrones de creciente complejidad. La primera capa puede identificar características básicas como picos en frecuencias específicas, capas intermedias detectan combinaciones como la relación entre armónicos, y capas profundas reconocen patrones complejos que corresponden a fallos específicos.

Lo que hace especial a este enfoque es su capacidad para detectar patrones sutiles que podrían pasar desapercibidos para el ojo humano. Puede identificar, por ejemplo, que cierta modulación en bandas laterales alrededor de la frecuencia de engrane, combinada con ligeros cambios en el factor de cresta, indica desgaste temprano en engranajes, incluso cuando todas las amplitudes individuales están dentro de rangos normales.

Ventajas:

- Capacidad de modelar relaciones no lineales complejas
- Robustez al ruido en señales de vibración
- Integración de múltiples fuentes de datos

Riesgos:

- Alto costo computacional
- Riesgo de sobreajuste requiere regularización
- Dificultad de interpretación (caja negra)
- Sensibilidad a inicialización de parámetros

Preparación de datos:

- Normalización estandarizada de características
- Codificación one-hot para variables categóricas
- Balanceo de clases mediante oversampling/weights
- División estricta train/validation/test

Máquinas de Vectores Soporte (SVM)

Las SVM funcionan encontrando el mejor límite de separación entre diferentes estados de la máquina en un espacio multidimensional. Imagina que cada medición de vibración es un punto en un espacio donde las coordenadas son todas las características medidas: RMS, amplitud 1X, amplitud 2X, relación 2X/1X, etc.

La SVM busca el hiperplano óptimo que mejor separe, por ejemplo, los puntos correspondientes a "operación normal" de los de "desbalanceo". Lo hace maximizando el margen entre las clases, creando una zona de seguridad donde pequeñas variaciones en las mediciones no cambian el diagnóstico.

Cuando los datos no son linealmente separables - como suele ocurrir en problemas reales de vibraciones - las SVM usan "kernels" para transformar los datos a un espacio de mayor dimensión donde sí se pueden separar. Esto permite capturar relaciones complejas no lineales, como el hecho de que la relación entre la amplitud fundamental y los armónicos puede cambiar de manera no lineal con la carga operativa.

Para problemas multiclase como el diagnóstico de vibraciones donde hay múltiples fallos posibles, las SVM usan estrategias como "uno contra uno." "uno contra todos", construyendo clasificadores binarios para cada par de fallos y combinando sus resultados.

Ventajas:

- Efectivo en espacios de alta dimensionalidad
- Teoría matemática sólida y márgenes óptimos
- Buen rendimiento con datos limitados

Riesgos:

- Costo computacional $O(n^2)$ en entrenamiento
- Sensibilidad a elección de kernel y parámetros
- Dificultad con datasets muy grandes

Preparación de datos:

- Escalado estandarizado obligatorio
- Codificación numérica de categóricas
- Imputación cuidadosa de valores faltantes
- Reducción de dimensionalidad para eficiencia

Cuadro 1: Comparativa de modelos en diagnóstico de vibraciones

Modelo	Estudio	Contexto	Métrica	Tamaño
PSO-BP	Zhang & Su (2010)	Diagnóstico turbogenerador	0.15 % error	10 fallos
XGBoost	Wang et al. (2021)	Clasificación turbinas	95.3 % F1	15k muestras
Random Forest	Zhang et al. (2020)	Rodamientos	92 % accuracy	2.4k muestras
SVM-RBF	Widodo & Yang (2007)	Fallos mecánicos	93.7 % accuracy	1.2k muestras

Conclusión

Cada modelo tiene fortalezas particulares para diferentes aspectos del problema:

Los árboles y bosques aleatorios son ideales para la interpretabilidad - puedes mostrar exactamente las reglas que llevaron a un diagnóstico, lo que es valioso para ganar la confianza de los ingenieros de planta.

XGBoost ofrece el mejor equilibrio entre precisión y eficiencia computacional, siendo capaz de manejar grandes volúmenes de datos en tiempo real mientras mantiene alta precisión diagnóstica.

Las redes neuronales son superiores para detectar patrones complejos y sutiles en los espectros de frecuencia, especialmente cuando se combinan múltiples fuentes de datos.

Las SVM son efectivas cuando se tienen conjuntos de datos más pequeños pero bien etiquetados, y son particularmente robustas frente a sobreajuste.

En la práctica, muchos sistemas implementan ensembles que combinan las fortalezas de múltiples approaches, usando cada modelo donde mejor se desempeña y combinando sus diagnósticos para mayor confiabilidad.

Aspectos Éticos y Cumplimiento

Mitigación de Sesgos

- Análisis de fairness por tipo de turbina
- Balanceo de datos por condición operativa
- Auditoría regular de decisiones

Explicabilidad

- SHAP analysis para importancia global y local
- Reportes interpretables para ingenieros
- Límites de confianza para predicciones

Privacidad y Seguridad

- Anonimización de identificadores
- Control de acceso a datos sensibles

- Cumplimiento ISO 13373-3 para trazabilidad