

RESUMEN

El aprendizaje automático (ML por sus siglas en inglés) es conocido como la rama de la inteligencia artificial que reúne algoritmos estadísticos, probabilísticos y de optimización, que aprenden empíricamente. ML puede aprovechar el conocimiento y la experiencia que se han generado durante años en las empresas para realizar automáticamente diferentes procesos. Por lo tanto, ML se ha aplicado a diversas áreas de investigación, que estudian desde la medicina hasta la ingeniería del software.

De hecho, en el campo de la ingeniería del software, el mantenimiento y la evolución de un sistema abarca hasta un 80% de la vida útil del sistema. Las empresas, que se han dedicado al desarrollo de sistemas software durante muchos años, han acumulado grandes cantidades de conocimiento y experiencia. Por lo tanto, ML resulta una solución atractiva para reducir sus costos de mantenimiento aprovechando los recursos acumulados. Específicamente, la Recuperación de Enlaces de Trazabilidad, la Localización de Errores y la Ubicación de Características se encuentran entre las tareas más comunes y relevantes para realizar el mantenimiento de productos software. Para abordar estas tareas, los investigadores han propuesto diferentes enfoques. Sin embargo, la mayoría de las investigaciones se centran en métodos tradicionales, como la indexación semántica latente, que no explota los recursos recopilados. Además, la mayoría de las investigaciones se enfocan en el código, descuidando otros artefactos de software como son los modelos.

En esta tesis, presentamos un enfoque basado en ML para la recuperación de fragmentos en modelos (FRAME). El objetivo de este enfoque es recuperar el fragmento del modelo que realiza mejor una consulta específica. Esto permite a los ingenieros recuperar el fragmento que necesita ser trazado, reparado o ubicado para el mantenimiento del software. Específicamente, FRAME combina la computación evolutiva y las técnicas ML.

En FRAME, un algoritmo evolutivo es guiado por ML para extraer de manera eficaz distintos fragmentos de un modelo. Estos fragmentos son posteriormente evaluados mediante técnicas ML. Para aprender a evaluarlos, las técnicas ML aprovechan el conocimiento (fragmentos recuperados de modelos) y la experiencia que las empresas han generado durante años. Basándose en lo aprendido, las técnicas ML determinan qué fragmento del modelo realiza mejor

una consulta. Sin embargo, la mayoría de las técnicas ML no pueden entender los fragmentos de los modelos. Por lo tanto, antes de aplicar las técnicas ML, el enfoque propuesto codifica los fragmentos a través de una codificación ontológica y evolutiva. En resumen, FRAME está diseñado para extraer fragmentos de un modelo, codificarlos y evaluar cuál realiza mejor una consulta específica.

El enfoque ha sido evaluado a partir de un caso real proporcionado por nuestro socio industrial (CAF, un proveedor internacional de soluciones ferroviarias). Además, sus resultados han sido comparados con los resultados de los enfoques más comunes y recientes. Los resultados muestran que FRAME obtuvo los mejores resultados para la mayoría de los indicadores de rendimiento, proporcionando un valor medio de precisión igual a 59.91%, un valor medio de exhaustividad igual a 78.95%, un valor-F medio igual a 62.50% y un MCC (Coeficiente de Correlación Matthews) medio igual a 0.64. Aprovechando los fragmentos recuperados de los modelos, FRAME es menos sensible al conocimiento tácito y al desajuste de vocabulario que los enfoques basados en información semántica. Sin embargo, FRAME está limitado por la disponibilidad de fragmentos recuperados para llevar a cabo el aprendizaje automático. Esta tesis presenta una discusión más amplia de estos aspectos así como el análisis estadístico de los resultados, que evalúa la magnitud de la mejora en comparación con los otros enfoques.