



Universidad de La Habana
Facultad de Matemática y Computación

Asignatura: Aprendizaje Automático

Meta-Learning

Integrantes

Jabel Resendiz Aguirre
Amalia Beatriz Valiente Hinojosa
Noel Pérez Calvo
Melanie Forsythe Matos
Jorge Alejandro Echevarría Brunet
Arianna Camila Palancar Ochando

Carrera: Ciencia de la Computación

Índice

1. Introducción al Meta-Learning	2
2.	4
3. Experimentación	4

1. Introducción al Meta-Learning

El meta-learning, o “aprendizaje a aprender”, es un área del aprendizaje automático que busca desarrollar algoritmos capaces de **aprender de experiencias pasadas para mejorar el rendimiento en nuevas tareas**. A diferencia del aprendizaje tradicional, donde un modelo se entrena para una tarea específica, el meta-learning se centra en **extraer conocimiento generalizable** que pueda transferirse a problemas desconocidos, acelerando el proceso de aprendizaje y mejorando la eficiencia.

En el contexto de AutoML (Automated Machine Learning), el meta-learning se utiliza para **seleccionar automáticamente algoritmos y configuraciones de hiperparámetros** adecuadas para un conjunto de datos dado, basándose en información obtenida de datasets anteriores. Esto permite reducir significativamente el tiempo de experimentación y evitar la necesidad de un ajuste manual exhaustivo de los modelos.

Conceptos Clave: Para comprender meta-learning y el enfoque que se presenta en este reporte, es importante familiarizarse con los siguientes conceptos:

- **Meta-features:** Son características que describen datasets. Por ejemplo, el número de instancias, el número de atributos, la distribución de las clases o medidas estadísticas de los atributos. Las meta-features permiten comparar datasets y predecir qué algoritmos o configuraciones funcionarán mejor.
- **Pipeline de Machine Learning:** Es la secuencia de pasos que se aplican a un dataset, desde la preparación de datos (preprocesamiento, normalización) hasta el entrenamiento y evaluación de un modelo. Cada etapa puede tener múltiples opciones e **hiperparámetros** que afectan el rendimiento final.
- **Hiperparámetros:** Son parámetros de un algoritmo de ML que no se aprenden directamente a partir de los datos, sino que deben fijarse antes del entrenamiento. Ejemplos incluyen la profundidad máxima de un árbol de decisión, la tasa de aprendizaje de un modelo de redes neuronales o el número de vecinos en un k-NN. La correcta elección de hiperparámetros es crucial para el desempeño de un modelo.
- **Tareas de aprendizaje:** Se refiere a un problema específico de aprendizaje automático que se desea resolver, definido por un dataset y un objetivo de predicción (por ejemplo, clasificación o regresión). En meta-learning, cada tarea puede considerarse como una instancia en la que el sistema debe seleccionar un algoritmo y sus hiperparámetros adecuados. El aprendizaje meta busca **transferir conocimiento entre tareas** para mejorar la eficiencia en tareas nuevas.
- **AutoML:** Se refiere a la automatización del proceso de selección de algoritmos y ajuste de hiperparámetros. Los sistemas AutoML tradicionales requieren probar múltiples configuraciones y evaluar su desempeño, lo que puede ser costoso en tiempo y recursos.

Motivación y Problema El problema central que se aborda con meta-learning es “cómo lograr que un sistema AutoML pueda predecir de manera eficiente qué algoritmos y configuraciones funcionarán bien en un nuevo dataset”, sin tener que evaluar exhaustivamente todas las posibilidades.

Aunque existen sistemas como AutoSkLearn, PMF u OBOE que implementan estrategias de selección y optimización, estos enfoques requieren lanzar experimentos para cada nuevo dataset (fase cold-start), lo que limita la eficiencia. Además, el diseño manual de meta-features no siempre garantiza una buena generalización entre datasets.

Por esta razón, surge la necesidad de aprender **meta-features que capturen la relación entre datasets y configuraciones óptimas de hiperparámetros**. Estas meta-features permiten:

- Establecer una topología confiable del espacio de datasets, donde datasets cercanos comparten configuraciones de hiperparámetros similares.
- Reducir costos computacionales al usar configuraciones de datasets “vecinos” en nuevos datasets.
- Obtener información interpretable sobre cuándo un algoritmo funciona bien, proporcionando intuición sobre la naturaleza del problema.

Descripción del Modelo: El modelo usado es un sistema de meta-learning diseñado para aprender automáticamente meta-features representativas de los datasets, las cuales se utilizan para guiar la selección de algoritmos y configuraciones de hiperparámetros en sistemas de AutoML. La idea central de MetaFeatX es construir una representación de los datasets que capture de manera efectiva la relación entre las características del conjunto de datos y las configuraciones de hiperparámetros que producen los mejores resultados. Para lograr esto, el sistema considera dos tipos de representaciones. La primera, llamada representación básica, consiste en un conjunto de meta-features manualmente diseñadas que describen estadísticamente los atributos del dataset, sus dimensiones y la distribución de clases. La segunda, denominada representación objetivo, describe cada dataset como la distribución de configuraciones de hiperparámetros que permiten alcanzar el mejor rendimiento; esta representación solo está disponible para datasets ya evaluados.

El aprendizaje de las meta-features en MetaFeatX se realiza mediante un procedimiento de Transporte Óptimo, que permite encontrar una transformación de las meta-features básicas hacia nuevas meta-features aprendidas. El objetivo de esta transformación es que la distancia euclíadiana entre datasets en el espacio de meta-features aprendidas emule la distancia Wasserstein-Gromov entre las distribuciones de configuraciones óptimas de los datasets. De esta forma, datasets que son similares según MetaFeatX tienden a compartir configuraciones de hiperparámetros similares, lo que permite un AutoML más eficiente y con menor costo computacional, evitando la necesidad de evaluar exhaustivamente todas las posibles configuraciones para cada nuevo dataset.

Entre las ventajas de MetaFeatX se encuentran la inicialización eficiente de sistemas AutoML en nuevos datasets, la reducción de la experimentación costosa, y la posibilidad de obtener interpretabilidad sobre qué características de un dataset influyen en el desempeño de un algoritmo. Adicionalmente, MetaFeatX permite estimar la dimensionalidad intrínseca del espacio de datasets respecto a un algoritmo de aprendizaje, proporcionando información sobre la complejidad relativa de los problemas. En comparación con sistemas AutoML clásicos, que aprenden un modelo de rendimiento por dataset individual, MetaFeatX se destaca por transferir conocimiento de datasets previos, caracterizándose como un enfoque de meta-learning. Esta transferencia de conocimiento permite evitar la fase de inicio en frío para nuevos datasets, mejorar el desempeño promedio en la selección de algoritmos y configuraciones, y generar representaciones generalizables que reflejan la topología del espacio de datasets de manera más eficiente.

2.

3. Experimentación