

Обзор GOLEM

Владимир Латыпов

30-01-2024

История появления

Изначально была библиотека **FEDOT** для AutoML, основана на pipeline-ах ~произвольной структуры (dag), поиск происходит посредством эволюции. Но алгоритм графовой оптимизации оказался полезен и для кучи других задач, в т.ч. проектов лаборатории:

- BAMT (Bayesian AutoML Tool)
- NAS (Neural Architecture Search)
- GEFEST (Generative Evolution For Encoded STructures)
- пользовательские применения (коллаборация с химической лабораторией, например — btw полезный подход)

Поэтому было решено выделить эту часть в отдельную библиотеку — GOLEM.

Глобально про алгоритм

Тюнинг

Многокритериальная оптимизация

Операторы

- Мутация
- Кроссовер
 - ↳ subtree — выбрать
- Селекция
- «Регуляризация»

Ближе к коду

Язык: Python. Библиотеки:

- `joblib + multiprocessing`
- `torch + karateclub` для контекстуального бандита на GNN
- Package `core` contains the main classes and scripts.
- Package `core.adapter` is responsible for transformation between domain graphs and internal graph representation used by optimisers.
- Package `core.dag` contains classes and algorithms for representation and processing of graphs.

- Package `core.optimisers` contains graph optimisers and all related classes (like those representing fitness, individuals, populations, etc.), including optimization history.
- Package `core.optimisers.genetic` contains genetic (also called evolutionary) graph optimiser and operators (mutation, selection, and so on).
- Package `core.utilities` contains utilities and data structures used by other modules.
- Package `serializers` contains class `Serializer` with required facilities, and is responsible for serialization of project classes (graphs, optimization history, and everything related).

- Package visualisation contains classes that allow to visualise optimization history, graphs, and certain plots useful for analysis.
- Package examples includes several use-cases where you can start to discover how the framework works.
- All unit and integration tests are contained in the test directory.

The sources of the documentation are in the docs directory.

Issues

Направления развития

NOTEARS

Theorem 1. *A matrix $W \in \mathbb{R}^{d \times d}$ is a DAG if and only if*

$$h(W) = \text{tr} \left(e^{W \circ W} \right) - d = 0, \quad (7)$$

where \circ is the Hadamard product and e^A is the matrix exponential of A . Moreover, $h(W)$ has a simple gradient

$$\nabla h(W) = \left(e^{W \circ W} \right)^T \circ 2W, \quad (8)$$

and satisfies all of the desiderata (a)-(d).

→ сводим к задаче непрерывной оптимизации; целевая функция и ограничения — дифференцируемы. Решаем с помощью метода расширенной Лагранжианы.

Ограничения:

- Целевая функция должна естественно (дифференцируемо, поменьше константа Липшица, легко вычислима) продолжаться на вещественные веса
- Как задать на пространстве, содержащем категориальные переменные?

Метаэволюция

Коэволюция

см. Признаки

Поддержка разнообразия

Expressive encodings