

Raport AutoPyTorch

4 maja 2022

1 Wstęp

Jestem członkiem zespołu ToJaDa i wybrałem zrobienie code review dla pakietu autoML o nazwie AutoPyTorch. Opracowany ten framework został przez zespół WTF.

2 Czy ten kod osiąga cel, który postawiono?

Celem każdego opracowania frameworku, było stworzenie prostej w obsłudze funkcji, która ma za zadanie na prawie każdym zbiorze (binarnym), przeprowadzić znalezienie jak najlepszego modelu. Sprowadzało się to więc na:

1. wczytanie zbiorów dla benchmarków
2. przeprowadzenie preprocessingu
3. stworzenie modelu

2.1 wczytanie zbiorów dla benchmarków

Stworzyli do tego autorzy notatnik `get_binary_tasks.ipynb`, który spełnia tę funkcję w 100%.

2.2 przeprowadzenie preprocessingu

Gdyż AutoPyTorch robi preprocessing, jedyne co stworzyli twórcy to zamiana typu kolumn z 'object' na 'category'.

2.3 stworzenie modelu

Notatnik `run.ipynb` dokładnie to robi, wraz z minimalnym preprocessingiem.

3 Czy w kodzie są jakieś oczywiste błędy logiczne?

Po moim dwukrotnym przejrzaniu kodu, nie udało mi się znaleźć żadnych błędów, tym bardziej oczywistych. Jednak jak wiadomo, sprawdzając po kimś wykonaną robotę, kiedy się nie ma klucza odpowiedzi, potrafi być mylące. W związku z tym najlepiej przejrzeć resztę raportów, czy na pewno nikomu też się nie udało nic znaleźć.

4 Czy patrząc na wymagania zawarte podczas prezentacji są one w pełni zaimplementowane?

Z tego co zapamiętałem, to twórcy nie skarżyli się na brak jakiejś implementacji, jak i ja nie potrafiłem znaleźć żadnej takiej wzmianki w slajdach prezentacji. Jeżeli natomiast chodzi o cztery poprzednie punkty, to powtórzę, że udało się wykonać wszystko.

5 Czy kod jest zgodny z istniejącymi wytycznymi stylistycznymi? (czy kod jest zgodny z PEP 8)

Nazewnictwo obiektów, czyli np. `'moja_funkcja'` zamiast `'mojaFunkcja'` oraz innych obiektów jest zgodna z PEP 8. Również, każdy cel obiektu jest znany za pomocą nazwy i/lub kontekstu. Komentarze gdzie są potrzebne również się znajdują. Jedynie do czego jestem w stanie się przyczepić, to do `'wyniki.ipynb'`, a dokładnie kilku ostatnich komórek notatnika, gdzie są bardzo długie linijki, a powinny zostać rozbite na kilka mniejszych.

6 Czy są jakieś obszary, w których kod mógłby zostać poprawiony? (skrócić, przyspieszyć, itp.)

Jedna rzecz którą bym zmienił (jeżeli miałbym ochotę, zdaję sobie sprawę, że jest to niekoniecznie potrzebne, gdyż nie ma co się męczyć z automatyzacją 30 minut jak można coś zrobić ręcznie w 5 minut) to zapisywanie wyników w `'wyniki.ipynb'`, aby nie było 22 zmiennych, a np. słownik posiadający te wyniki jako swoje wartości a klucze jako nazwy zbiorów, albo zwykła lista posiadająca same wyniki. Ewentualnie jeszcze można by się pokusić o wychwytywanie i obsługę wyjątków, ale gdyż jest to `.ipynb` nie jest to konieczne.

7 Czy dokumentacja i komentarze są wystarczające?

Moim zdaniem tak. Nie ma użytych bardzo skomplikowanych funkcji ani zawiłych. Zdecydowanie wszystko można było się albo domyślić (po nazwie, kontekście) albo przeczytać z komentarzy. Natomiast należy wziąć pod uwagę, że sam już trochę programuję, więc możliwe, że dla kogoś początkującego nie wszystko, choć nadal większość, byłaby zrozumiała.

8 Czy udało się odtworzyć zamieszone przykłady w kodzie?

Niestety, kod pobierający zbiory danych nie spełnił swojego zadania i nie pobrał wszystkich zbiorów, w czas który mu na to wyznaczyłem. Możliwe, że jednak te 30 minut to był za mały czas dla mojego internetu. W związku z tym, uważam, że trzeba wziąć pod uwagę inne review tego kody, gdyż prawdopodobnie była to tylko i wyłącznie moja wina. W związku z tym, że nie udało mi się pobrać danych w przystępnym czasie, nie sprawdzałem reszty kodu.

9 Podsumowanie

Podsumowując, kod mi się bardzo podobał jak i slajdy prezentacji. Nie udało mi się znaleźć czegoś, co by budziło we mnie wątpliwości albo odrażenie do kodu :) Mam nadzieję, że to czego nie udało mi się sprawdzić uda się ocenić za pomocą innych code review odnośnie pracy tego zespołu.