

MMDS Challenge

Raport

Imię Nazwisko Indeks
Imię Nazwisko Indeks

15 stycznia 2019

1 Wprowadzenie

Raport powinien stanowić podsumowanie działań podjętych w celu stworzenia modelu predykcyjnego do konkursu Data Ninja. Powinien on zawierać opis podjętych działań w celu przetworzenia danych, stworzenia ich reprezentacji, wybory, wytrenowania i ulepszania modelu, oraz analizę uzyskanych wyników.

2 Wymagania

Raportu powinien zawierać następujące sekcje:

2.1 Sformułowanie problemu

Krótki i konkretny opis **zadania** konkursowego, z analizą miary, według której oceniane są rozwiązania w konkursie.

2.2 Opis metody predykcyjnej

Opis modelu zaprojektowanego w celu wykonania zadania konkursowego. Należy opisać model jaki został użyty, przedstawić proces jego uczenia i doboru hiperparametrów. Osiągnięte wyniki należy odnieść do wyników uzyskanych z użyciem bardzo prostego rozwiązania, stanowiącego punkt odniesienia dla wyników. Może być on prostym modelem OVR dla najpopularniejszych etykiet (z zawartą najlepszą osiągalną wartością miary oceny oraz osiągniętą jej wartością przez model).

Model konkursowy ma stanowić próbę uzyskania lepszych wyników niż model bazowy. Należy wybrać jeden lub kilka adekwatnych modeli, przeprowadzić dla nich poprawnie przygotowanie danych, optymalizację parametrów uczenia oraz walidację wyników.

Należy zwrócić szczególną uwagę na jakość metodyki przeprowadzenia doboru parametrów i walidacji wyników. Wyniki zaproponowanej metody należy odnieść do wyników osiągniętych przez rozwiązanie bazowe.

Powinny również zostać poruszone kwestie wstępnego przetwarzania danych czy ekstrakcji cech, ewentualnej inżynierii cech podjętej w celu uzyskania lepszej reprezentacji danych. Mogą się pojawić tutaj kwestie związane ze stemmingiem czy lematyzacją, czy transformacjami cech. Należy wspomnieć jaki format danych jest odpowiedni dla wybranego algorytmu uczenia modelu.

2.3 Wyniki eksperymentalne

W tym punkcie należy umieścić i opisać wyniki uzyskane w konkursie. Należy umieścić wyniki uzyskiwane podczas pracy nad rozwiązaniem na danych treningowych i walidacyjnych przy doborze parametrów algorytmów uczących. Należy zamieścić komentarze oraz tabele z wynikami (Tabela 1 jest przykładem jak stworzyć tabelę w $\text{\LaTeX}$ ’u).

W raporcie należy zamieścić wartości miary jakości:

- na użytym zbiorze treningowym (wartości błędu minimalizowanego podczas treningu i konkursowej miary oceny)
- na użytym zbiorze walidacyjnym (wartości błędu minimalizowanego podczas treningu i konkursowej miary oceny)
- na zbiorze testowym (wartość konkursowej miary oceny, wg sprawdzarki na stronie konkursu).

Można również zamieścić i skomentować czasy przetwarzania danych, uczenia i predykcji.

Należy podać jaka miara jest optymalizowana podczas treningu modelu, jeśli jest taka dostępna.

Do utworzenia modelu predykcyjnego użytego do stworzenia predykcji dla zbioru testowego można użyć wszystkich danych dostępnych w konkursie (nie tylko wyodrębnionego zbioru użytego do treningu po oddzieleniu danych użytych do walidacji).

Wynik na zbiorze testowym (ze sprawdzarki) może być podany dla części wybranych modeli. Należy zweryfikować czy wyniki uzyskiwane podczas walidacji odpowiadają wynikom na zbiorze testowym na sprawdzarce.

Należy wskazać jakie parametry/dane zostały użyte do wytrenowania każdego z modeli dla których podawane są wyniki.

2.4 Podsumowanie

Na końcu jest zawsze miejsce na krótkie podsumowanie.

	zbiór treningowy					zbiór walidacyjny					test
	L	P@1	P@3	P@5	AP	L	P@1	P@3	P@5	AP	AP
model1											
model2											
model3											

Tablica 1: Przykładowa tabela z wynikami eksperymentalnymi. L oznacza miarę optymalizowaną podczas treningu wybranego modelu, a AP to $AvgPrec@1, 3, 5$.

3 Uwagi

Raport powinien zawierać krótki opis podjętych działań i przeprowadzonych badań. Nie należy opisywać użytych narzędzi czy użytych metod, chyba, że są to metody specjalnie zaprojektowane dla rozwiązania tego zadania.

Należy przytoczyć ich nazwę, najlepiej także anglojęzyczną aby uniknąć niespójności wynikających z różnych tłumaczeń.

Opis rozwiązania powinien być na tyle szczegółowy by można było na jego podstawie zreprodukować uzyskane wyniki.

Należy skomentować wyniki, a nie jedynie zamieścić liczby w raporcie.

4 Punktacja

Za raport można uzyskać do 50 punktów. Oceniane będą:

- staranność i sensowność działań podjętych w celu przetworzenia danych, stworzenia rozwiązania i jego ulepszenia,
- jakość analizy wyników i wniosków.

Na ocenę z raportu nie wpływa miejsce zajęte w rankingu na stronie konkursu.

Całość raportu nie może przekraczać 10 stron.