

Rozpoznawanie kierowców na podstawie profilu jazdy

Jastrzębski Mateusz, Gadawski Łukasz

1 czerwca 2015

Abstract if need

1 Treść zadania

Realizacja projektu polega na stworzeniu rozwiązania zadania konkursowego dotyczącego rozpoznawania kierowców na podstawie podanych danych. Organizator konkursu dostarcza profile jazdy dla 3613 kierowców. Dla każdego kierowcy dane są 200 plików CSV opisujących pojedynczą podróż. Każdy taki plik zawiera współrzędne kierowcy na płaszczyźnie w każdej sekundzie jego podróży, np:

```
x,y
0.0,0.0
18.6,-11.1
36.1,-21.9
...
```

W każdym zbiorze 200 podróży zamieszczono kilka losowych tras innych kierowców (niekoniecznie z dostarczonego zbioru).

Zadaniem projektu jest identyfikacja podróży, które nie należą do danego kierowcy. Dane wynikowe powinny zostać przedstawione jako tablica wartości prawdopodobieństwa, że konkretna ścieżka należy do aktualnego kierowcy, np:

```
driver_trip,prob
1_1,0.8
1_2,0.4
1_3,0.7
1_4,0.9
...
```

2 Wykorzystane algorytmy

2.1 Random Forest

Celem metod zespołowych (ang. *ensemble methods*) jest połączenie prognozy estymatorów wraz z odpowiednim algorytmem uczącym, aby uzyskać możliwość oceny pojedynczej próbki w jakim stopniu jest zgodna z innymi próbkami na podstawie wspomnianych estymatorów. Do tego celu wyróżnia się dwa rodzaje metod :

- metody uśrednianiające (ang. *averaging methods*) - polegające na stworzeniu estymatorów niezależnie, a następnie na uśrednieniu ich prognoz, np. Random Forest,
- metody wzmacniające (ang. *boosting methods*) - polegające na sekwencyjnej budowie estymatorów

Jednym z wykorzystanych algorytm jest algorytm Random Forest [cite]. W projekcie została wykorzystana implementacja w języku python z pakietu scikit-learn [cite]. Jego praktyczne wykorzystanie wiąże się z następującymi krokami:

1. inicjalizacja modelu,
2. dostarczenie danych,
3. predykcja nowych.