

Zadanie domowe nr 3

Temat do wyboru

Rzeczy organizacyjne

Maks. punktów: **7 pkt**

Termin oddania: email ze sprawozdaniem do **18 stycznia (piątek), do godz. 23:59** (chciałbym mieć dzień przed zajęciami na sprawdzenie państwa sprawozdań), omówienie projektu będzie 20 stycznia. Treść emaila:

Tytuł: PDZ-3-wybor, Jan Kowalski, nr indeksu: 123456
Treść: pusta
Załączniki: sprawozdanie.pdf
(+ skrypt.txt w przypadku wyboru trzeciego tematu)

Tematy

Proszę wybrać interesujący Państwa temat (jeden z poniższych), samodzielnie go opracować przeprowadzając testy i zamieścić szczegóły w sprawozdaniu w formacie PDF. Tematy nie są szczegółowo sformułowane, dają Państwu dość dużo luzu w wyborze baz danych czy technik.

1. Przetwarzanie obrazu

Wybierz interesujący Cię zbiór danych obrazków np. ze strony Kaggle (<https://www.kaggle.com/datasets>) lub UCI Machine Learning Repository (<https://archive.ics.uci.edu/ml/datasets.html>).
Zbiory danych z obrazkami (po filtrze): [Kaggle Image Datasets](#)

Celem projektu jest dokonanie klasyfikacji na wybranym zbiorze. Należy przetestować kilka klasyfikatorów (min. 3), które nadają się do przetwarzania obrazów. Kilka propozycji – można wybrać inne:

- Konwolucyjne Sieci Neuronowe (CNN), być może w kilku odmianach – google „Deep Learning for Image Classification”
- Random Forest lub Gradient boosting
- Support Vector Machines (SVM)

W sprawozdaniu: opisz krótko każdy z zastosowanych klasyfikatorów, wklej komendy z jego stosowania oraz wyniki tj. dokładność i macierz błędów. Na końcu napisz krótkie podsumowanie.

Alternatywnie: zamiast klasyfikacji obrazów, możesz wybrać inny interesujący Cię temat pod warunkiem, że dotyczy przetwarzania obrazów, data mining i sztucznej inteligencji.

2. Przetwarzanie tekstu

Należy znaleźć bazy danych z dłuższymi fragmentami tekstu, wiele ciekawych jest na Kaggle: [Kaggle Text Mining](#) czy [UCI Image Repository](#) . Następnie trzeba na danej bazie zastosować różne techniki związane ze zgłębianiem tekstu. Np.

- Preprocessing: usuwanie stop-words, usuwanie odmian słów, tworzenie macierzy DTM.
- Klasyfikacja tekstów (o ile pozwala na to baza danych). Przykładowo jeśli w bazie danych jest kolumna z tekstem i kolumna z klasą: czy można zgadnąć wartość klasy na podstawie tekstu? Jakie klasyfikatory mają tutaj sens?
- Sentiment Analysis: czy tekst jest pozytywny czy negatywny? Jakie inne emocje zawarte są w treści? Czy da się je rozszyfrować na podstawie występujących w nim słów?

Alternatywnie: można samodzielnie stworzyć bazę danych i przeprowadzić na niej eksperymenty: Przykładowo:

- Pobieram kilka tysięcy postów z Facebooka wraz z ikonami emocji (ich liczba związana z danym postem)



Pytania: jakie posty? Z jakiej strony? Jak je pobrać (Facebook Api, Selenium czy inne web-scrapery)? Proszę się zastanowić.

- Dokonuję przetwarzania i obróbki.
- Dokonuję Sentiment Analysis na zbiorze. Jakie posty są uważane za wesołe? Jakie za smutne? Jakie słowa kluczowe? Czy da się wysnuć jakieś socjologiczne czy psychologiczne wnioski na temat emocji i tematyki postów?

3. Sieci neuronowe sterujące czołgami

Na studiach dziennych jednym z projektów było napisanie sieci neuronowej do sterowania czołgiem. Mogą Państwo zdecydować się na wybór tego projektu:

<https://inf.ug.edu.pl/~gmadejsk/IO2017/PD-3-tank.pdf>

Sposób oceniania może się zmienić w zależności od tego ile dostanę rozwiązań.

4. Time Series Analysis

Przetestuj kilka technik analizy szeregów czasowych na wybranej bazie danych. Opisz po krótku każdą z technik. Zawrzyj komendy / fragmenty kodu z eksperymentów.