

Komputerowe systemy rozpoznawania

2020/2021

Projekt 1. Klasyfikacja dokumentów tekstowych

Autor Pierwszy, nr albumu 1

Autor Drugi, nr albumu 2

Prowadzący: tytuł/stopień imię nazwisko

5 lutego 2024

Opis projektu ma formę artykułu naukowego lub raportu z zadania badawczego (doświadczalnego/obliczeniowego) (wg indywidualnych potrzeb związanych np. z pracą inżynierską/naukową/zawodową). Kolejne sekcje muszą być numerowane i zatytułowane. Wzory są numerowane, tablice są numerowane i podpisane nad tablicą, rysunki są numerowane i podpisane pod rysunkiem. Podpis rysunku i tabeli musi być wyczerpujący (nie ogólnikowy), aby czytelnik nie musiał sięgać do tekstu, aby go zrozumieć.

Wybrane sekcje (rozdziały sprawozdania) są uzupełniane wg wymagań w opisie Projektu 1. i Harmonogramie Zajęć na WIKAMP KSR jako efekty zadań w poszczególnych tygodniach.

1 Cel projektu

Zwięzły (2-3 zdania) opis problemu badawczego/obliczeniowego, uwzględniający część badawczą i implementacyjną. **Nie przepisuj literatury ani teorii – napisz krótko jak rozumiesz to co masz wykonać: jakie działania, na jakim zbiorze danych (link lub przypis), jaki jest spodziewany efekt.**

Zamieszczony opis (własny, nie skopiowany) zawiera przypisy do literatury (bibliografii) zamieszczonej na końcu raportu/sprawozdania zgodnie z Polską Normą cytowania bibliografii (zob. materiały BG PŁ pt. „Bibliografia załącznikowa”).

Sekcja uzupełniona jako efekt zadania Tydzień 02 wg Harmonogramu Zajęć na WIKAMP KSR.

2 Klasyfikacja nadzorowana metodą k -NN. Ekstrakcja cech, wektory cech

Krótki opis metody k -NN: zasada działania, wymagane parametry wejściowe, format i znaczenie wyników/rezultatów. Opis własny z przypisami do literatury – minimum teorii potrzebnej do zadania, tak by inżynier innej specjalności zrozumiał dalszy opis [1]. **Nie przepisuj literatury ani teorii – napisz krótko jak rozumiesz to co masz wykonać w tym konkretnym przypadku.**

Wyekstrahowane cechy liczbowe i tekstowe dokumentów, min. 10 cech, w tym min. 2 o wartościach tekstowych, wszystkie opisane słownie oraz wzorami, z objaśnieniem oznaczeń i przykładami użycia, do tego precyzyjny opis możliwych wartości, które przyjmuje dana cecha (ułatwiający czytelnikowi zrozumienie znaczenia w zadaniu klasyfikacji). Pamiętaj, że wybrane cechy muszą reprezentować obiekt niezależnie od innych tekstów w tym samym lub w innym zbiorze. Podaj postać wektora wartości cech po procesie ekstrakcji. Użyte oznaczenia są jednolite w całym raporcie/sprawozdaniu.

Sekcja uzupełniona jako efekt zadania Tydzień 02 wg Harmonogramu Zajęć na WIKAMP KSR.

3 Miary jakości klasyfikacji

Miary jakości klasyfikacji (Accuracy, Precision, Recall, F1). We wprowadzeniu zaprezentować minimum teorii potrzebnej do realizacji zadania, tak by inżynier innej specjalności zrozumiał dalszy opis. Należy podać **konkretne wzory miar użyte w tym eksperymencie oraz krótko opisać ich znaczenie i zakresy przyjmowanych wartości. Należy podać przykładowe wartości każdej miary. Nie przepisuj teorii, ale podaj link/przypis i opisz jak rozumiesz jej zastosowanie w tym konkretnym zadaniu.**

Stosowane wzory, oznaczenia z objaśnieniami znaczenia symboli użytych w doświadczeniu. Oznaczenia jednolite w obrębie całego sprawozdania. Opis zawiera przypisy do bibliografii zgodnie z Polską Normą, (zob. materiały BG PŁ). **Sekcja uzupełniona jako efekt zadania Tydzień 03 wg Harmonogramu Zajęć na WIKAMP KSR.**

4 Metryki i miary podobieństwa tekstów w klasyfikacji

Wzory, znaczenia i opisy symboli zastosowanych metryk z przykładami. Wzory, opisy i znaczenia miar podobieństwa tekstów zastosowanych w obliczaniu metryk dla wektorów cech z przykładami dla każdej miary [2]. Oznaczenia jednolite w obrębie całego sprawozdania. **Podaj metryki i miary podobieństwa nie z literatury (te wystarczy zacytować linkiem), ale konkretne ich**

postaci stosowane w zadaniu. Jakie zakresy wartości przyjmują te miary i metryki, co oznaczają ich wartości? Podaj przykładowe wartości dla przykładowych wektorów cech.

Sekcja uzupełniona jako efekt zadania Tydzień 04 wg Harmonogramu Zajęć na WIKAMP KSR.

5 Wyniki klasyfikacji dla różnych parametrów wejściowych

Wstępne wyniki miary Accuracy dla próbnych klasyfikacji na ograniczonym zbiorze tekstów (podać parametry i kryteria wyboru wg punktów 3.-8. z opisu Projektu 1.). **Sekcja uzupełniona jako efekt zadania Tydzień 05 wg Harmonogramu Zajęć na WIKAMP KSR.**

6 Dyskusja, wnioski, sprawozdanie końcowe

Wyniki kolejnych eksperymentów wg punktów 2.-8. opisu projektu 1. Każdorazowo podane parametry, dla których przeprowadzana eksperyment. Wykresy (np. słupowe) i tabele wyników obowiązkowe, dokładnie opisane w „captions” (tytułach), konieczny opis osi i jednostek wykresów oraz kolumn i wierszy tabel.

****Ewentualne wyniki realizacji punktu 9. opisu Projektu 1., czyli „na ocenę 5.0” i ich porównanie do wyników z części obowiązkowej**.** Dokładne interpretacje uzyskanych wyników w zależności od parametrów klasyfikacji opisanych w punktach 3.-8 opisu Projektu 1. Szczególnie istotne są wnioski o charakterze uniwersalnym, istotne dla podobnych zadań. Omówić i wyjaśnić napotkane problemy (jeśli były). Każdy wniosek/problem powinien mieć poparcie w przeprowadzonych eksperymentach (odwołania do konkretnych wyników: wykresów, tabel). Dla końcowej oceny jest to najważniejsza sekcja sprawozdania, gdyż prezentuje poziom zrozumienia rozwiązywanego problemu.

****** Możliwości kontynuacji prac w obszarze systemów rozpoznawania, zwłaszcza w kontekście pracy inżynierskiej, magisterskiej, naukowej, itp. ******

Sekcja uzupełniona jako efekt zadań Tydzień 05 i Tydzień 06 wg Harmonogramu Zajęć na WIKAMP KSR.

7 Braki w realizacji projektu 1.

Wymienić wg opisu Projektu 1. wszystkie niezrealizowane obowiązkowe elementy projektu, ewentualnie podać merytoryczne (ale nie czasowe) przyczyny tych braków.

Bibliografia

- [1] R. Tadeusiewicz: Rozpoznawanie obrazów, PWN, Warszawa, 1991.
- [2] A. Niewiadomski, Methods for the Linguistic Summarization of Data: Applications of Fuzzy Sets and Their Extensions, Akademicka Oficyna Wydawnicza EXIT, Warszawa, 2008.

Literatura zawiera wyłącznie źródła recenzowane i/lub o potwierdzonej wiarygodności, możliwe do weryfikacji i cytowane w sprawozdaniu.