

Modelowanie i analiza zjawisk obiegu informacji w wirtualnej infosferze on-line z elementami m.in. technik web mining, text mining, big data, analizy sentymentu, i elementami głębokiego uczenia maszynowego.

Marcin Sydow
Instytut Podstaw Informatyki PAN, Warszawa
msyd@ipipan.waw.pl

1. Opis projektu.

Obieg informacji w wirtualnej infosferze (media społecznościowe, blogi, tweety, wiadomości on-line, komentarze pod nimi, wyniki wyszukiwarek, etc.) istotnie wpływa na zachowanie już nie tylko poszczególnych jednostek, ale całych społeczeństw. Potwierdzone w ostatnich latach wydarzenia np. prób zwiększania polaryzacji społeczeństwa, wpływu na wyniki wyborów demokratycznych (np. w USA) lub katalizowania masowych protestów społecznych (np. tzw. “arabska wiosna”), kampanii dezinformacyjnych (np. tzw. “polskie obozy zagłady”, etc.) przyczyniły się do używania terminu “wojny informacyjnej”. Istnieje pilna potrzeba obiektywnych nowatorskich badań w/w zjawisk i procesów z zastosowaniem najnowszych technologii informacyjnych w tym np. technik typu word-embedding, głębokie sieci neuronowe (np. sieci rekurencyjne typu LSTM i ich warianty), słowniki wydźwięku, etc. Celem projektu jest studiowanie w/w problemów, aktualnego stanu wiedzy i rozwój nowych modeli, metod i algorytmów dedykowanych badaniu i modelowaniu takich zjawisk, nie tylko dla języka angielskiego, ale szczególnie dla infosfery dotyczącej języka polskiego.

Badania w obszarach:

- a. gromadzenie danych nieustrukturyzowanych i ustrukturyzowanych (np. tekst, dane grafowe, artykuły wiadomości, blogi, komentarze, tweety, etc.) dostępnych on-line w sieci
- b. analiza w/w danych i modelowanie pod względem m.in.: tematyki, wydźwięku, wpływie na opinię publiczną, manipulacji informacją, tzw. “fake newsów”, prób masowego wpływania np. Na wyniki wyborów, zjawisk tzw. “trollingu”, etc.

Cele pracy m.in. obejmują:

- a. intensywne studia obecnego stanu wiedzy (literatura) i ewaluacja dotychczasowych modeli
- b. projektowanie i implementacja nowych i ulepszonych metod i modeli
- c. stworzenie zasobów (m.in. zbiorów danych) wspomagających realizację danego zadania
- d. stworzenie metod i zasobów ewaluacji metod
- e. uzyskanie ciekawych wyników eksperymentów empirycznych na w/w danych
- f. wysokiej jakości publikacje naukowe opisujące w/w wyniki

2. Wymagania

- a. Ukończone studia drugiego stopnia z informatyki, matematyki, fizyki lub równoważna wiedza i gotowość do samodzielnego uzupełnienia wiedzy, umiejętności programistyczne i znajomość podstawowych technik przetwarzania danych
- b. Duża samodzielność i determinacja
- c. Podstawowa wiedza dotycząca uczenia maszynowego, statystyki, algorytmów i eksploracji danych
- d. Zainteresowanie współczesnymi problemami obiegu informacji w społeczeństwie, zainteresowanie technikami data science, w tym web mining i analizą tekstu (text mining) i rozwiązywaniem rzeczywistych problemów, mile widziana dobra orientacja w bieżących zagadnieniach geopolitycznych, społecznych i gospodarczych
- e. Dobra znajomość języka angielskiego (przede wszystkim czytanie i pisanie)
- f. Chęć uczestniczenia w regularnym przygotowywaniu wysokiej jakości publikacji naukowych

Przykładowa Literatura:

- 1) Ronald E. Robertson, Shan Jiang, Kenneth Joseph, Lisa Friedland, David Lazer, and Christo Wilson. 2018. Auditing Partisan Audience Bias within Google Search. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 148 (November 2018), 22 pages. DOI: <https://doi.org/10.1145/3274417>
- 2) Eberl, J.-M., Boomgaarden, H. G., & Wagner, M. (2017). One Bias Fits All? Three Types of Media Bias and Their Effects on Party Preferences. *Communication Research*, 44(8), 1125–1148. <https://doi.org/10.1177/0093650215614364>
- 3) Shan Jiang and Christo Wilson. 2018. Linguistic Signals under Misinformation and Fact-Checking: Evidence from User Comments on Social Media. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 82 (November 2018), 23 pages. DOI: <https://doi.org/10.1145/3274351>