

Modeling and analysis of information flow in on-line virtual info-sphere with elements of web mining, text mining, big data, sentiment analysis and deep learning.

Marcin Sydow
Institute of Computer Science PAS, Warsaw
msyd@ipipan.waw.pl

1. Project description

Information flow in the on-line virtual info-sphere (social media, blogs, messaging services, on-line news and comments below them, search engine results, etc.) crucially influences the opinion, stance and behaviour of the particular citizens but also whole societies. Some recently confirmed cases of attempts to increase social/political opinion polarization, influence/manipulate the process of democratic elections (e.g. USA) or supporting massive social protests (e.g. “Arab spring”), massive misinformation/defamation campaigns, contributed to increased popularity of the phrase “info war”. There is an urgent need to start and carry high-quality objective and unbiased research concerning the phenomenons mentioned above, with the application of the modern information technologies, including web/text mining, word-embedding, sentiment analysis, deep learning (e.g. LSTM or other recurrent deep neural networks), etc.

The goal of the project is to study the state-of-the-art in this area and developing new, improved models, methods and algorithms for the above-mentioned research domain, including the ones targetted at Polish language.

Research in the following areas:

- a. collecting semi-structured and non-structured (text, graphs, on-line news articles, blogs, comments, tweets, etc.) data available on-line, etc.
- b. analysis of the above information and modeling with regard to (for example): topics, sentiment, public opinion influence, information manipulation, fake news, election results manipulation, “trolling”, etc.

The goals include:

- a. intensive research of current state-of-the-art (literature) and evaluation of the existing models
- b. design and implementation of improved methods and models
- c. creation of resources (e.g. data sets) supporting the realisation of the research
- d. design and implementation of the evaluation methods
- e. interesting results of intensive experimental research
- f. high-quality academic research publications (top international conferences and journals)

2. Requirements

- a. M.Sc. of Computer Science, Mathematics, Physics or equivalent/comparable knowledge, eagerness to independent self-studying; programming and data manipulation skills
- b. Independence and high motivation
- c. Basic knowledge concerning machine learning, statistics, algorithms and data mining.
- d. Interest in contemporary problems of virtual information flow in society, interest in data science (web/text mining, etc.), in solving real practical problems; interest in contemporary geopolitics, social and economy issues, etc.
- e. Good Polish and English skills
- f. Motivation and eagerness to regularly publish high-quality academic research papers

References (examples):

- 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>