A Multi-Modal approach for FAke News discovery and propagation from big Data ANalysis and artificial intelliGence Operations

DAVID MARTÍN-GUTIÉRREZ[1*], GUSTAVO HERNÁNDEZ-PEÑALOZA[2*], JOSE MANUEL MENÉNDEZ[3*], AND FEDERICO ÁLVAREZ[4*]

[*] *Visual Telecommunication Applications Group. Signals, Systems and Radio communications Department, SSR, Universidad Politécnica de Madrid*
[1] *dmz@gatv.ssr.upm.es*
[2] *ghp@gatv.ssr.upm.es*
[3] *jmm@gatv.ssr.upm.es*
[4] *fag@gatv.ssr.upm.es*

The analysis of data collected from diverse sources such as news, claims, open data and media can provide valuable insights for early detection and propagation traceability of events related to fake news. This article describes a novel system that involves data ingestion, processing and analysis to extract relevant features and serve as a means of verification of multimedia contents. The modular ecosystem treats the information from a modality perspective (i.e. text, images, videos, meta-data) and offers both single and global evaluation of the content's trustworthiness probability. The system is being implemented and tested in relevant topic scenarios including migration, climate change and Europe.

1. INTRODUCTION

The "post truth" era in politics and in public opinion shaping is so broadly accepted that there is a lemma in the Oxford Dictionary[1], that it also includes the following quota: "In this era of post-truth politics, it's easy to cherry-pick data and come to whatever conclusion you desire". A survey conducted by the Pew Research Centre towards the end of last year found that 64% of social networks users said made-up news stories were causing confusion about the basic facts of current issues and events[2]. Fake News are now a hot issue in Europe as well as worldwide, particularly referred to Political and Social Challenges that reflect in business as well as in industry. Working out who to trust and who not to believe has been a facet of human history since our ancestors began living in complex societies. The main difference in today's societies is that technology and the Internet give us access to a vast amounts of information. Most of the times even more information than we can ingest. Although, the internet enabled many voices to be heard, which otherwise would not, our concern now is how to control and filter-out the dissemination of things that seem to be untrue. Europe is lacking a systematic knowledge and data transfer across organizations to address the aggressive emergence of the well-known problem of fake news and post-truth effect. The possibility to use cross sector Big Data management and analytics, along with an effective interoperability scheme for all our data sources, will generate new business and societal impacts involving several stakeholders and targets such as:

i Media Companies (Press agencies, news, TV Broadcaster etc)

ii Governmental institutions and organisations

iii The overall Industrial ecosystem

iv The entire society

The main and critical challenge is to establish the reliability of online information. This challenge must be faced from multiple perspectives: *(a)* European tradition in democracy, journalism and transparency should play a worldwide example in fast changing society, where all citizens appears completely overwhelmed by the new technologies and by the new social challenges. *(b)* Having a large number of people in a society who are misinformed is absolutely devastating and extremely difficult to cope with. *(c)* News Agencies and the other media companies are the first target and operator that use and disseminate news. In the past, news and media companies accessed data as separate "silos" without a strong interoperability between different media. This practice enforced the possibilities for malicious media sources to disseminate fake news.

In this paper, a modular approach to tackle the mentioned issues is presented. This system provides a set of modules to analyse information from multiple sources to support users (i.e.

[1] Oxford Dictionary Post truth definition, see https://en.oxforddictionaries.com/definition/post-truth

[2] http://www.journalism.org/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/

journalists) in the verification process. These modules rely on the use Artificial Intelligence (AI) methods to detect trustworthiness or manipulation indicators. In this paper, the overall architecture, data model and work-flow are described.

2. SYSTEM ARCHITECTURE OVERVIEW

The proposed approach is based on three main modalities: text, multimedia content such as videos and images and finally, metadata content. More specifically, depending on the modality, one or more micro-services were developed to extract relevant features that may lead to promote the detection of fake news. Thus, the following services are considered:

i A Text Analysis service based on Natural Language Processing by considering the article body.

ii A Topic Extraction service based on Named Entity Recognition (NER) techniques by analysing the article body.

iii A Multimedia analysis service based on image and video analysis.

iv A Source-credibility service based on graph analytics techniques.

In addition, two micro-services were implemented as well to both collect and pre-process the information from the different data silos. These services have a crucial role in the process since the data is totally unstructured depending on the different data silos and therefore, it needs to be mapped into a unique data model to be used by the rest of the analysers.

Moreover, an efficient procedure for ingesting large volumes of articles is required to retrieve the information needed for each of the aforementioned services. To do so, Big Data tools such as Elasticsearch, Kafka or Apache Nifi were considered when building the final architecture. Figure 1 represents the Big Data architecture proposed by FANDANGO to assess both Fake News detection and content manipulation. More specifically, FANDANGO has already collected more than 11 million of articles from different countries such as Italy, Spain, Belgium, Greece, United Kingdom and United States.

Furthermore, the ingestion process starts by a Crawling service via Apache Nifi which analyses several websites to extract meaningful information such as the headline, the body, the authors as well as the multimedia content among other features. As Figure 1 shows, every service produces data into a specific Kafka topic and it consumes data from another. Subsequently, the preprocessing service reads from Kafka the set of crawled documents and it starts both cleaning and organising the information to adequate the input of the different analysers that later on will perform the classification task. More specifically, each of the aforementioned analysers provides the system with a score which indicates the level of trustworthiness regarding the modality.

Finally, a fusion model is employed to obtain a unique score that provides an overall indicator of the level of trustworthiness of the article based on the different modalities considered during the analysis stage.

3. MULTIMODAL ANALYSIS FOR FAKE NEWS DETECTION

Once the documents are properly pre-processed, the remaining analyser services access to the specific information according to its modality.

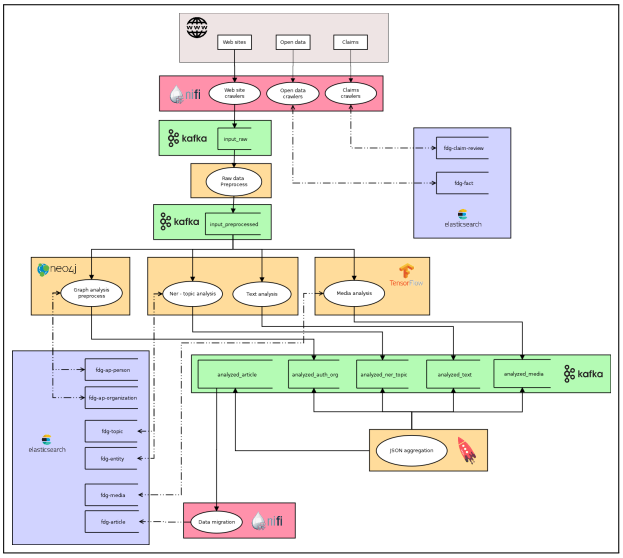

Fig. 1. General architecture of FANDANGO system.

More specifically, the text analyser service gets the content body of the document, analyses it via advanced Natural Language Processing (NLP) procedures and provides the system with a score of trustworthiness regarding the article content. Furthermore, the topic extractor service performs a NER procedure [1] to retrieve the main topics of the document using its content as well. Regarding the scores, both text and topic analysers compute the trustworthiness indicator for the article content itself.

On the other hand, the multimedia analyser service retrieves the set of images and videos associated to the document and attempts to detect manipulations in the content using powerful neural networks such as DeepFake [2, 3]. Thus, the score of this service is related to the multimedia content of the document.

The remaining source-credibility service collects the information related to publishers, authors, articles and topics and generates a graph to connect the different entities involved in the process. In particular, this service provides a trustworthiness indicator for both publisher and authors by computing a set of centrality methods [4] to measure the impact of each node in the network together with a set of metrics based on the connection of the nodes to indicate the polarity of the impact (level of trustworthiness).

Finally, considering the set of scores for each service, a multimodal model based on advanced neural networks is employed to fuse all the indicators and thus, to provide journalists and press agencies with an overall metric of how reliable a certain article is according to all the components that are involved in its creation.

4. FANDANGO USER INTERFACE

As a breakthrough of the system, a User Interface has been developed in order to encourage media professionals such as journalists to use FANDANGO as a potential tool to verify news using the different services that were described above. Moreover, the interface allows end-users to navigate throughout the different services and review the features that each particular service has used to perform the decision of its score. By doing so, end-users may analysed the performance of the AI modules and thus, their confidence and commitment in using the platform to

this purpose increase.

A. Platform Validation

The whole platform of FANDANGO has been tested and validated by different end-users throughout Pilots which has fostered promising results regarding time consuming and confidence. More specifically, it was observed that the time spent by professionals when verifying the credibility of a certain article was drastically reduced by using the platform. Moreover, the confidence of the results provided by the different analysers were investigated as well in order to verify the performance of the different algorithms.

5. GENERAL CONCLUSIONS

This study has presented a multi-modal solution for analysing large volumes of media content in order to detect potential fake news providers. Moreover, FANDANGO provides a complete multimedia-based Dataset with more than 11 million of articles which can be used for improving the proposed architecture in future investigations.

ACKNOWLEDGMENT

This work was supported by the H2020 European Project: FANDANGO. Grant no. 780355.

REFERENCES

1. J.-H. Wang, T.-H. Su, and K.-T. Chen, "Exploiting multi-document term extraction to improve named entity recognition for major concept detection," in *2019 International Conference on Technologies and Applications of Artificial Intelligence (TAAI),* (IEEE, 2019), pp. 1–6.
2. B. Dolhansky, R. Howes, B. Pflaum, N. Baram, and C. C. Ferrer, "The deepfake detection challenge (dfdc) preview dataset," arXiv preprint arXiv:1910.08854 (2019).
3. I. Amerini, L. Galteri, R. Caldelli, and A. Del Bimbo, "Deepfake video detection through optical flow based cnn," in *Proceedings of the IEEE International Conference on Computer Vision Workshops,* (2019), pp. 0–0.
4. T. Agryzkov, L. Tortosa, J. F. Vicent, and R. Wilson, "A centrality measure for urban networks based on the eigenvector centrality concept," Environ. Plan. B: Urban Anal. City Sci. **46**, 668–689 (2019).