Computational Fact Checking is Real, but will it Stop Misinformation?*
Invited Talk - Extended Abstract

Paolo Papotti

EURECOM, France
papotti@eurecom.fr

Abstract. Fact checkers are overwhelmed by the amount of false content that is produced online every day. To support fact checking, several research efforts have been focusing on automatic verification methods to assess claims and experimental results show that such methods enable effective labeling of textual content. However, while fact checkers start to adopt some of these tools, the misinformation fight is far from being won. In this talk, I cover the opportunities and limitations of computational fact checking and its role in fighting misinformation.

Fake news and misinformation in social media is a real societal problem. As social media keep being used to mislead communities, there have been more and efforts for manual and automated fact checking [6]. In this talk, I first talk about the big push in both fact checking research and industry that has been observed after the start of "infodemic" crises related to the COVID-19 pandemic. Indeed, to support human fact checkers, a lot of research has been done to cover several manual steps in the time-consuming process of verifying claims. I focus on the computational verification of a textual claim, therefore assuming that worth checking claims have already been identified from a given text.

Automatic verification in general is hard, but there are promising results for different kinds of claims. For example, claims about numerical values, such as "In March 2020, confirmed case of COVID in Italy increased by 900%", can be verified automatically with access to official statistics [5]. This is an example of a *reference approach*, where the claim is checked against a trusted source [3, 4] Success here clearly depends on the availability and quality of the available data, which can vary widely across topics. Promising results are also reported for property claims, especially for popular entities [1, 2]. Most of these methods are based on learning algorithms trained on existing annotated claim corpora [8]. These *machine learning approaches* use probabilistic models to predict whether a given claim is likely to be correct. Finally, *contextual approaches* use metadata

to estimate the factuality of a claim, for example, by analysing the language used, who is spreading or repeating a claim, and if it has already been verified [7, 9].

New algorithms keep increasing the accuracy and the scope of the computational verification approaches, but several issues prevent their adoption in fact checking organizations. One important problem is that most real claims are complex and subtle and go beyond the reach of the automatic solutions. For example, a claim stating "COVID vaccines have been adopted despite incomplete testing campaign over humans" require fact checkers to collaborate with domain experts, collect and check multiple pieces of evidence, and write a rebuttal that is aware of the context and the framing of the claim. While algorithms can help in this process, the final verification (and explanation of such decision) requires a human understanding beyond current AI solutions.

Finally, I conclude with observations about the next steps that we need to take to improve the coverage and effectiveness of computational fact checking. We definitely need to keep pushing the technical research. In this agenda, one of the most important points is to focus on interpretability of the proposed solutions. Algorithms and models should output explainable checking decisions, with considerations about possible bias in the reference information. However, technical advancement is not going to be effective in reducing the misinformation problem unless a real holistic approach is taken to tackle it. Misinformation is a societal issue that involves actors ranging from politicians to tech companies, with duties that are still to be defined clearly. As computer science researches, we have the responsibility to design and deploy the best algorithms and models to attack the problem, but also the obligation to make clear that misinformation is not a crisis that will be solved only by developing more advanced AI solutions.

References

1. Ahmadi, N., Lee, J., Papotti, P., Saeed, M.: Explainable fact checking with probabilistic answer set programming. In: TTO (2019)
2. Augenstein, I., Lioma, C., Wang, D., Chaves Lima, L., Hansen, C., Hansen, C., Simonsen, J.G.: MultiFC: A real-world multi-domain dataset for evidence-based fact checking of claims. In: EMNLP (2019)
3. Chen, W., Wang, H., Chen, J., Zhang, Y., Wang, H., Li, S., Zhou, X., Wang, W.Y.: TabFact: A large-scale dataset for table-based fact verification. In: ICLR (2020)
4. Huynh, V.P., Papotti, P.: A benchmark for fact checking algorithms built on knowledge bases. In: CIKM (2019)
5. Karagiannis, G., Saeed, M., Papotti, P., Trummer, I.: Scrutinizer: A mixed-initiative approach to large-scale, data-driven claim verification. VLDB **13**(11) (2020)
6. Nakov, P., Corney, D.P.A., Hasanain, M., Alam, F., Elsayed, T., Barrón-Cedeño, A., Papotti, P., Shaar, S., Martino, G.D.S.: Automated fact-checking for assisting human fact-checkers. In: IJCAI. pp. 4826–4832. ijcai.org (2021)
7. Shaar, S., Babulkov, N., Da San Martino, G., Nakov, P.: That is a known lie: Detecting previously fact-checked claims. In: ACL (2020)
8. Thorne, J., Vlachos, A., Cocarascu, O., Christodoulopoulos, C., Mittal, A.: The fact extraction and VERification shared task. In: FEVER (2018)

9. Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. Science **359**(6380) (2018)