Rahul Das
CS 4395.001
Dr. Karen Mazidi

Summary of "InfoSurgeon: Cross-Media Fine-grained Information Consistency Checking for Fake News Detection"

Author List: Yi R. Fung[1], Chris Thomas[2], Revanth Reddy[1], Sandeep Polisetty[3], Heng Ji[1], Shih-Fu Chang[2], Kathleen McKeown[2], Mohit Bansal[4], Avirup Sil[5]

[1] University of Illinois at Urbana-Champaign, [2] Columbia University, [3] UMass Amherst, [4] University of North Carolina at Chapel Hill, [5] IBM

Paper Citations on Google Scholar: 24

As the internet has become an integral part of modern society, the information on the internet has become increasingly influential. Like any method of control, misinformation on the internet has become the new frontier for bad actors to push false, chaotic, or dangerous agendas and opinions. While historically this was done by hand, as natural language processing is developing into an increasingly expansive and powerful field, much of the generation of fake news has been relegated to NLP neural networks. Since these models can generate misinformative content at an incredibly and increasingly rapid pace, it is important for there to be tools available that can also process and identify fake content quickly.

Current forms of misinformation detection are based on lexical features, semantic structure, and consistency with known data. The main issue that current systems run into is that there are different forms of misinformation. There can be direct lies and factual inconsistency, but the more common, and more effective, method is to mix misinformation in with correct information to make it appear more reliable. This can be significantly harder to detect since a smaller manipulation of real information might not be the kind of "simple" inconsistency that other systems look for, such as semantic inconsistency between article images and captions or blatant lies inconsistent with previously known data. While completely fabricated facts can be fairly reliably classified through methods such as detecting previously fact-checked claims (Shaar) or finding paths between concepts on knowledge graphs to validate factual relationships (Ciampaglia). Another traditional approach has been to detect stylistic characteristics that could indicate fake news (Perez-Rosas), but news does not have a standardized format which can make style-based detection difficult for models. The size of news can vary from small phrases to a mid-sized article with a couple paragraphs, up to large multimedia articles. Additionally, as trained generative models become better, the stylistic patterns that indicate fake news are

becoming more difficult to detect (Schuster). As fake news generation models improve and generate increasingly complex and natural articles, the detection methods used previously tend to fall short when it comes to reliable classification.

Misinformation can happen on a smaller scale and does not follow a reliable standard. To account for this possible variation in the nature of fake news, the researchers behind InfoSurgeon designed the model to validate all the knowledge elements of an article in the context of the full multimedia news article. By treating the whole article as one interconnected claim, even minor inconsistencies can be detected that might have slipped past other detection algorithms that are designed for detecting specific kinds of inconsistency. An Information Extraction system is used to extract the entities, relations, and events from the article. By using a graph based neural network along with an external knowledge base, InfoSurgeon builds a knowledge graph of the article and all of its media. Examining the relationships between the informational elements of the graph allows the model to detect inconsistencies that would not be found by existing models that focus on specific components or attributes of the article. Additionally, the researchers used the InfoSurgeon model to also generate fake news articles that have individual knowledge element manipulations. By generating and observing fake articles using a complex model, the researchers were better able to understand how articles can be faked with more complex manipulations than previous models are designed to detect. The focus on alterations to individual information elements is what differentiates InfoSurgeon from existing systems and improves on the techniques already established.

To evaluate the effectiveness of InfoSurgeon, researchers used two datasets. One dataset is made up of 32,000 real news articles from the NYTimes-NeuralNews dataset, an established benchmark dataset for fake news detection, as well as 32,000 fake articles generated by the Grover fake news generation model (Zellers). The second dataset used was the new VOA-KG2txt dataset, made up of 15,000 real articles from Voice of America, as well as 15,000 generated fake news articles. To compare InfoSurgeon's performance to previous models, the same datasets were processed by two recent baseline models. The model

Approach	NYTimes-NeuralNews	VOA-KG2txt
Zellers et al. (2019)	56.0%	86.4%
Tan et al. (2020)	77.6%	88.3%
InfoSurgeon	94.5%	92.1%

Table 1: A comparison of document-level misinformation detection accuracy on the two datasets.

from (Tan) was chosen because it is most similar to InfoSurgeon, but it does not use knowledge graphs. The model from (Zellers) was chosen because it is good for detecting Grover-based generated fake articles. The classification accuracies of the three models were measured on the two datasets and are shown here in Table 1. InfoSurgeon greatly outperformed the baseline systems on the same datasets due to its novel focus on elements of fake articles that previous models were not designed for. The knowledge graph structure of the model allowed InfoSurgeon to identify the semantics of the articles and any inconsistencies more easily, which makes it more receptive to a wider variety of fine-grained fake news articles. The researchers also evaluated InfoSurgeon's performance on the more complex knowledge element-level misinformation that existing systems struggle with. In Table 2, seen below, the performance of InfoSurgeon's element-level classification is compared to a random baseline classification system. The results indicate that InfoSurgeon's approach achieves a significantly higher accuracy than the baseline. The accuracy of 31%-37% may seem low but classifying misinformative elements of a text is a very complex task and requires high-level reasoning and extensive background knowledge. With the complexity of these classification tasks in mind, InfoSurgeon's performance, especially when contrasted with the baseline models, is quite impressive on both article and element-level classification.

Approach	VOA^0	VOA
Random	16.6%	16.9%
InfoSurgeon	36.5%	31.3%

Table 2: Knowledge element-level misinformation detection F-score on the VOA (*VOA-KG2txt*) dataset, consisting of entity swapping, link insertion, and subgraph replacement manipulations, and its easier variant, VOA^0, which contains entity swappings.

This paper was written by some very qualified researchers who have extensive careers performing research in computer science and machine learning. In total, the works of these researchers have been cited 121,118 times according to Google Scholar. The most cited of these researchers is Prof. Shih-Fu Chang from Columbia University, with 66,492 citations to his work. The InfoSurgeon model developed by these researchers is a significant step in the development of anti-fake news mechanisms. With such a high level of classification accuracy from the model, the framework provides a good resource for assessing news articles. The knowledge element-focused approach of the model also provides fine-grained explanations for why articles are classified as such. InfoSurgeon is a very interesting and powerful tool for combating misinformation in the world of quickly and easily manipulated news.

Rahul Das
CS 4395.001
Dr. Karen Mazidi

Paper Summarized

Yi Fung, Christopher Thomas, Revanth Gangi Reddy, Sandeep Polisetty, Heng Ji, Shih-Fu Chang, Kathleen McKeown, Mohit Bansal, and Avi Sil. 2021. InfoSurgeon: Cross-Media Fine-grained Information Consistency Checking for Fake News Detection. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages 1683–1698, Online. Association for Computational Linguistics.

Prior Works

Giovanni Luca Ciampaglia, Prashant Shiralkar, Luis M Rocha, Johan Bollen, Filippo Menczer, and Alessandro Flammini. 2015. Computational fact checking from knowledge networks. PloS one, 10(6):e0128193.Shaden Shaar, Nikolay Babulkov, Giovanni Da San Martino, and Preslav Nakov. 2020. That is a known lie: Detecting previously fact-checked claims. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3607–3618, Online. Association for Computational Linguistics.

Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. 2018. Automatic Detection of Fake News. In Proceedings of the 27th International Conference on Computational Linguistics, pages 3391–3401, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Tal Schuster, Roei Schuster, Darsh J Shah, and Regina Barzilay. 2020. The limitations of stylometry for detecting machine-generated fake news. *Computational Linguistics*, pages 1–12.

Rowan Zellers, Ari Holtzman, Hannah Rashkin, Yonatan Bisk, Ali Farhadi, Franziska Roesner, and Yejin Choi. 2019. Defending against neural fake news. In *Advances in Neural Information Processing Systems*, pages 9054–9065.