Chapter 3
Misinformation Containment Using NLP and Machine Learning:
Why the Problem Is Still Unsolved

Vishnu S. Pendyala
https://orcid.org/0000-0001-6494-7832
San Jose State University, USA

ABSTRACT

Despite the increased attention and substantial research into it claiming outstanding successes, the problem of misinformation containment has only been growing in the recent years with not many signs of respite. Misinformation is rapidly changing its latent characteristics and spreading vigorously in a multi-modal fashion, sometimes in a more damaging manner than viruses and other malicious programs on the internet. This chapter examines the existing research in natural language processing and machine learning to stop the spread of misinformation, analyzes why the research has not been practical enough to be incorporated into social media platforms, and provides future research directions. The state-of-the-art feature engineering, approaches, and algorithms used for the problem are expounded in the process.

INTRODUCTION

Social media has been subject to plenty of controversies owing to its use for spreading misinformation, sometimes to the extent of manipulating a country's presidential elections (Pendyala et al., 2018). The objective of this chapter is to explain some of the recent machine learning and natural language processing approaches for misinformation containment and provide reasons why, despite the large quantity of research in the area, the problem is still unsolved. Modeling domains using math has time and again proven to yield solutions to some of the toughest problems in the past. Machine learning, for the most part, has evolved from applied math. There has been an upsurge in the literature on the topic of trust in social media using machine learning models in recent times. This chapter starts with a survey of some

DOI: 10.4018/978-1-6684-6001-6.ch003

of the machine learning models, methods, and techniques that have been used to address the problem of the trustworthiness of the information on the Internet, which helps in misinformation containment.

The techniques are discussed under various sub-heads such as language models, few-shot learning, bot detection, graph theoretic approaches to misinformation containment, and using Generative Adversarial Network models for detecting fake multimedia content as well as textual content. As Table 1 shows, the corpus of articles on this topic is tremendous. A comprehensive survey of the existing literature is beyond the scope of this work. The survey is mainly intended to convey the underlying techniques and the resulting success that is reported in the literature and then to show why despite the claimed success, the problem is largely unsolved. The selection of the survey sub-topics in this chapter is based on the author's perception of what is indicative of the emerging literature.

As can be seen in the following sections, researchers have reported substantial success in misinformation containment (MC). However, even the layman can see that the problem is far from resolved. Information platforms such as WhatsApp have adopted means that are far from satisfactory to control the spread of lies on the Internet. For instance, by limiting the number of times a post can be forwarded, WhatsApp is curtailing useful information as well and not just malicious posts. Google search engine still returns web pages with a significant amount of misinformation and does not always indicate or quantify its belief in the fetched search results. Platforms such as Facebook depend on social media community standards to police the usage and are often a cause for grief for users who have genuine interests in posting information. Using formal methods such as First-Order-Logic can prove to be effective as well (Pendyala, 2018) but for focus and brevity, this chapter discusses only trends in machine learning and particularly in deep learning that seem promising. This chapter addresses the challenges in solving the misinformation containment problem and suggests some future directions.

BACKGROUND

Fake news continues to be a major problem. It is undoubtedly a complex problem to solve and appropriately attracted plenty of attention from the research community. A wide variety of machine learning algorithms such as support vector machines and logistic regression (Patel & Meehan, 2021), ensemble techniques like random forest (Antony Vijay et al., 2021) and Adaboost, deep learning frameworks such as LSTM (Rajalaxmi et al., 2022) and GAN (Xie et al., 2022), language models like BOW / TF-IDF (Mondal et al., 2022) and BERT (Palani et al., 2022), and many more have been tried out in the attempts to solve the problem. In terms of feature engineering as well, no stone has been left unturned. Manual feature extraction, graph embeddings (Karpov & Glazkova, 2020), and other approaches to representation learning (ElSherief et al., 2021) have all been tried. Not just supervised and unsupervised learning, but various other types of learning such as few-shot learning (Lo et al., 2022), meta-learning (Kozik & Chora's, 2022), transfer learning (Ghayoomi & Mousavian, 2022), meta-transfer learning (Shen, 2022), self-supervised learning (Huh et al., 2018), semi-supervised learning (Li et al., 2022), reinforcement learning (Mosallanezhad et al., 2022) (He et al., 2022), and active learning (Sahan et al., 2021) have been explored extensively for the problem. Figure 1 illustrates some of the approaches explored for misinformation containment. Despite the voluminous research literature purporting to solve the problem using machine learning methods, misinformation containment is largely unsolved and is growing by the day. The chapter provides some insights into the current state-of-the-art solutions and analyzes why they are not helping enough. The chapter will present some future directions that can help.