1-1-2023

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

Vishnu S. Pendyala
San Jose State University, vishnu.pendyala@sjsu.edu

Recommended Citation

Vishnu S. Pendyala. "Misinformation Containment Using NLP and Machine Learning: Why the Problem Is Still Unsolved" *Deep Learning Research Applications for Natural Language Processing* (2023): 41-56. https://doi.org/10.4018/978-1-6684-6001-6.ch003

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.

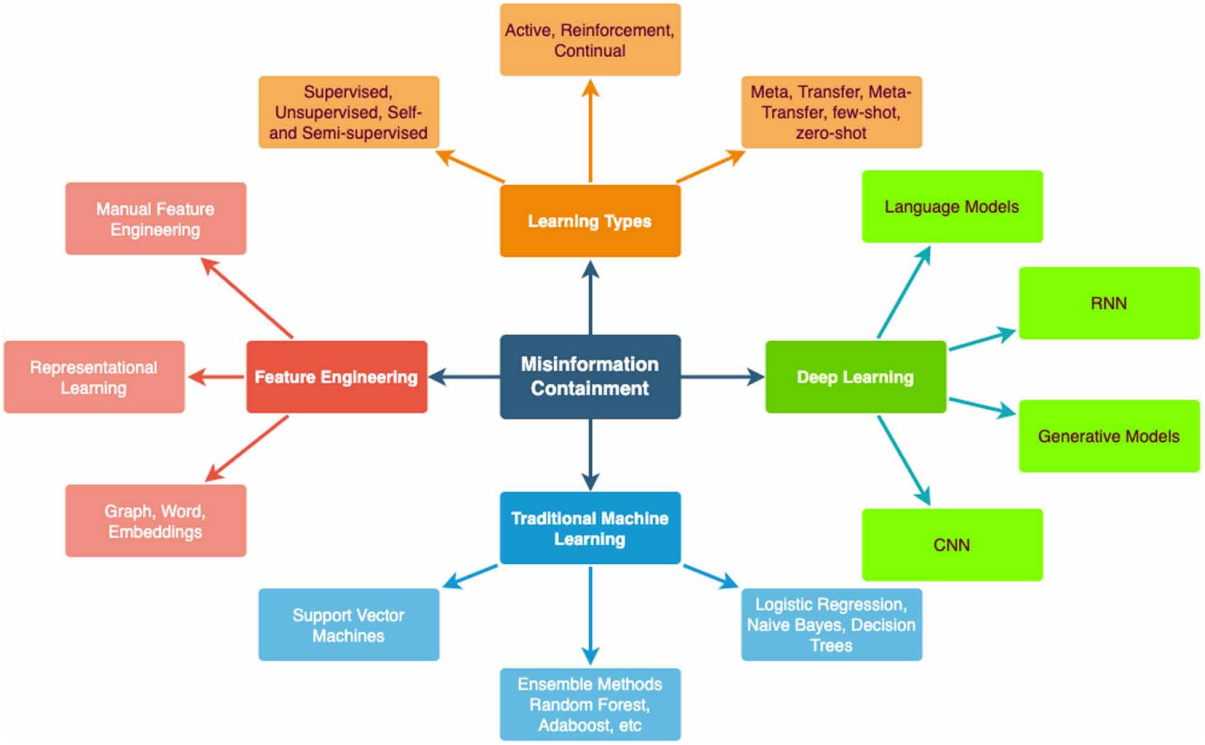
The techniques are discussed under various sub-heads such as language models, few-shot learning, bot detection, graph theoretic approaches to misinformation containment, and using Generative Adversarial Network models for detecting fake multimedia content as well as textual content. As Table 1 shows, 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.

Figure 1. Some of the approaches for Misinformation Containment that have been explored in the literature

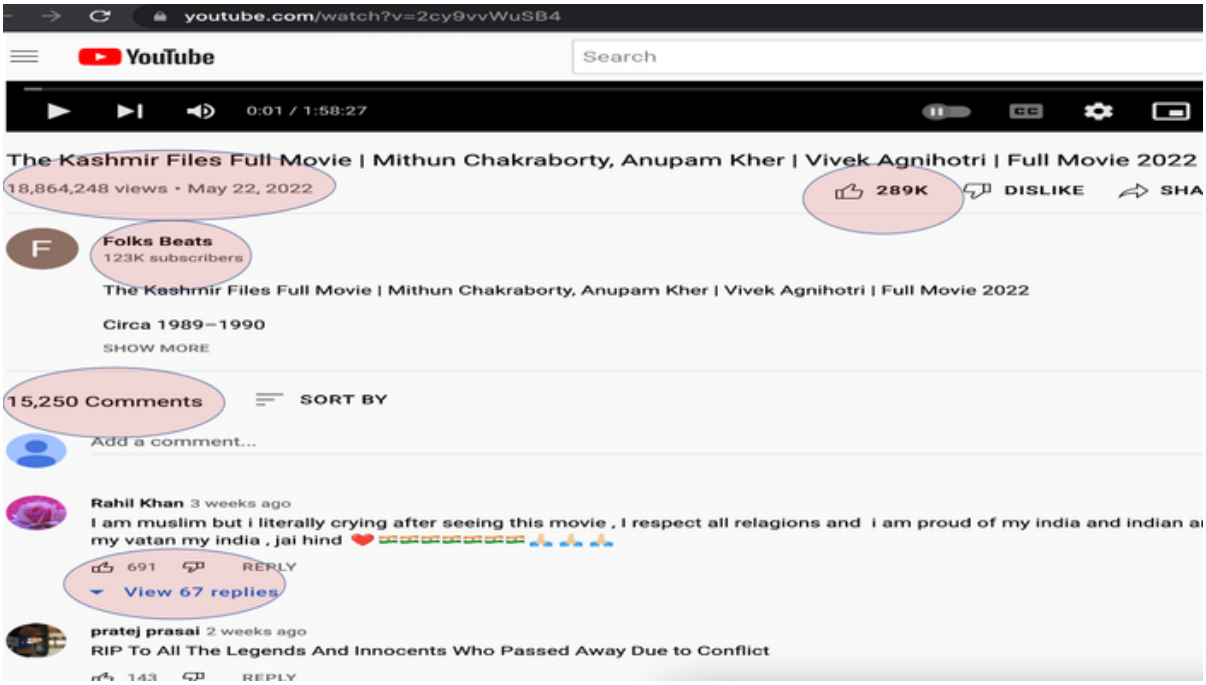


A Google search for "fake news" on July 23, 2022, returned 81.5 million results, including a fake news item relating to the Ukrainian President. The results also show that the Seattle times runs a section titled, "This week in fake news." To illustrate the problem further, Figure 1 is a screenshot of the metadata of a fraudulent upload on YouTube faking as the popular Indian movie, "Kashmir Files." It received tremendous attention from gullible viewers who seem to have believed that it is indeed the real movie. This is just one instance of how fake posts are largely uncontained.

Table 1. Google Scholar results listing articles purporting to solve the fake information problem

Search String	Article Count
"Machine learning" fake	103,000
"Deep learning" fake	48,800
"Language model" fake	6,830
"LSTM" fake	17,300
"graph neural networks" fake	17,800

Figure 2. A fake upload not matching the actors or attributes of the real movie gets ~19 million views, 289,000+ likes, and 15,250 genuine comments in 2 months (screenshot taken on July 23, 2022)

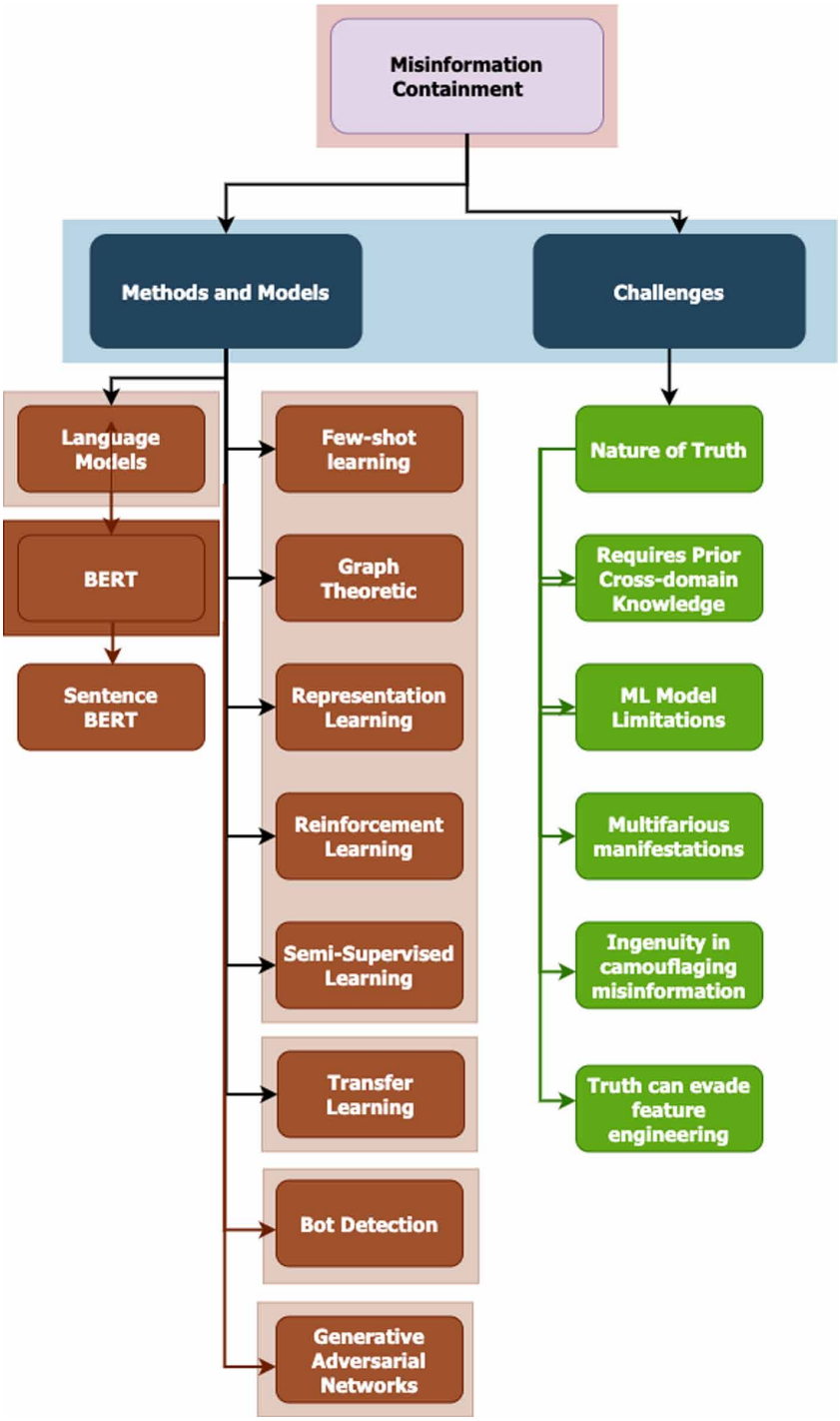


On the other hand, Table 1 shows the number of articles indexed by Google scholar that apparently use the popular technologies given in the search string to tackle the fake news problem. Depending on the underlying Google's search algorithm, the numbers may or may not be accurate, but the search results are quite indicative. The problem has attracted tremendous interest from the research community. Several articles that the author surveyed report outstanding successes. A "Fake News Challenge (FNC-1)[1]" was organized in 2017 to seek Machine Learning, Natural Language Processing, and Artificial Intelligence solutions to combat fake news and there are plans to organize FNC-2. The problem is still largely unsolved in the real world. It is therefore pertinent to research this huge disconnect between what is claimed in research and the actual reality.

Misinformation containment has been proven to be NP-hard more than a decade ago (Budak et al., 2011), which makes it a good candidate for approximate models such as the ones generated by machine learning. Given that artificial neural networks serve as universal function approximators, misinformation containment can be framed as a function of the features of the information that outputs whether the information is true or false or a degree of truthfulness. Deep learning that uses deep neural networks is rapidly expanding its scope of applications to the extent of prompting a debate in some corners as to whether traditional machine learning techniques are even relevant today. Deep learning generalizes better and performs better when the classification is nonlinear as in this case. Researchers (Wanda & Jie, 2020) used a deep learning architecture called Convolutional Neural Network (CNN) to detect fake profiles in online Social Networks. They achieve high accuracy in doing so. Like for any supervised machine learning classification problem, the authors extract features, train a model, handle overfitting issues by using regularization and apply the model to test data. They compare the results obtained using

the CNN architecture with the conventional machine learning models such as Logistic Regression and SVM and confirm that CNN performs better. We, therefore, focus more on deep learning approaches in the following sections. Figure 3 summarizes the flow of the rest of the chapter.

Figure 3. Misinformation Containment: Methods and challenges



Methods and Models

In the following paragraphs, although there is a reference to both deep learning and traditional machine learning-based approaches as they relate to the misinformation containment problem, most of the discourse is on deep learning. The purpose again is not for an exhaustive survey, which is almost infeasible given the amount of literature but to provide a high-level overview of some of the emerging trends not covered in other surveys on the topic.

Few-Shot Learning

The problem of misinformation detection has been addressed using meta-algorithms in traditional machine learning such as random forest and extra tree classifier with substantial reported success (Hakak, et al., 2021). The authors claim that the experiments resulted in 100% training and test accuracy on one dataset but training and testing accuracy of 99.8% and 44.15% respectively on a different training set (Hakak, et al., 2021). Few-shot learning (FSL) uses meta-learning that can work with fewer training examples. Model-agnostic meta-learning (MAML) is one such FSL technique that has been shown to result in better performance on fake news classification as compared to a host of other machine learning models (Salem, et al., 2021). Authors (Lwowski & Najafirad, 2020) propose identifying a latent space using self-supervised learning for few-shot learning and claim good results.

Language Models

Language models like BERT and Sentence BERT are quite popularly used for misinformation classification and efforts have been made to improve the classification performance by combining with enhanced attention-based methods (Paka, Bansal, Kaushik, Sengupta, & Chakraborty, 2021). When even human beings are not good at detecting fake news by just reading them, merely generating embeddings using NLP techniques is not sufficient for the task. Mining a wider corpus for additional features related to the information that needs to be classified can improve the results (Deepak & Chitturi, 2020). However, the secondary features so added are metadata like domain name and author details, which may not make a substantial difference. A similar approach is taken in (Braşoveanu & Andonie, 2019), but the augmented features include relations extracted from knowledge graphs. Despite the successes reported, language models based classification of misinformation in the form of short sentences is fundamentally flawed (Mifsud et al., 2021). It has been confirmed (Guderlei & Aßenmacher, 2020) that pre-trained BERT-based language models are good to start with for a subsequent transfer learning task for stance detection.

Graph Theoretic Approaches

Social media is often used synonymously with Online Social Networks. social media can be modeled as a network or a graph of users, and artifacts from the user posts. Such a model can then be used to predict the trust or credibility in the media. If we know the credibility of the nodes of a subset of this graph, that information can then be used by machine learning algorithms to estimate the credibility of the remaining graph. Researchers (O'Brien et al., 2019) prove that graph dependencies play an important role in credibility estimation in social media. We can relate this to the real world, where the credibility of a well-connected person can be much more easily established than a completely isolated person. The

idea can be related to the concept of homophily in psychology, where similar people are expected to bond with each other. Graph theoretic techniques can be used to model connections in social media and then exploit these homophily tendencies. The authors (O'Brien et al., 2019) use traditional machine learning algorithms such as Logistic Regression and Decision trees on the local and relational features based on the graphical structures, to achieve reasonably good accuracies in estimating credibility.

Learning from the neighborhood of a graph node using deep learning and generalizing the function learned to unseen nodes in the graph has many applications. Authors (Ghafari et al., 2019) use and extend Stanford University's GraphSAGE (Hamilton et al., 2017), which does exactly this to predict the trust between a pair of reviewers on an online review website such as Epinions. The authors use two datasets, one of which is from Epinions to do their experiments. This is an improvement over the Web-of-trust approach for Epinions like websites surveyed earlier (Pendyala, 2019) because it takes context into account. Trust is often contextual; an entity can be trusted in certain contexts but not all. As the authors (Ghafari et al., 2019) point out, most trust computing frameworks ignore this fact, whereas this work (Ghafari et al., 2019) leverages it. The system developed extracts contextual features from user demographics and reviews and uses them for the classification of the pairwise trust.

The work discussed above extracts a graphical structure from the social media entities and components to predict trustable relationships. Work has also been done the other way around to create a graphical framework in which entities are assured of trustable exchanges even in the presence of malicious players. Using several trust-based parameters and math around them, authors (Urena et al., 2020) create a robust framework for reputation-based communication. The authors run simulations to evaluate its performance and obtain good results. The Online Social Network (OSN) view of social media brings out the need for graph theoretic approaches for analyzing trust and credibility relationships between the entities in the social media space. We discussed only a few such approaches in this section, but until the problem of distrust is entirely solved, which probably is unlikely, we can expect to see a growth in the graph-theoretic-based approaches to address the problem.

Representation Learning

Manual feature engineering is increasingly getting replaced by representation learning. The goal of representation learning is to derive (or learn) a representation of the data automatically. The representation is usually in the form of an embedding, typically a vector. The embeddings can then be processed like any other feature vectors, possibly as the input layer for artificial neural networks. Representation learning is particularly effective with natural language and graphs. From the literature survey, we present a case study each, for natural language and graphs in the context of misinformation containment. Researchers (Borges et al., 2019) have used representation learning for stance detection as described in the "Fake News Challenge – 1 (FNC-1)" problem description1. Stance detection determines if two pieces of information agree with each other. In the FNC-1 case, the agreement is between the headline and the body of a news snippet. Using word embeddings from the pre-trained model, bi-directional RNNs such as LSTM and GRU, maxpooling, and attention mechanism, the researchers (Borges et al., 2019) model long news articles. Representation learning comes in handy when the data is a large graph. Social network graphs are large and evolving. New nodes can get added rapidly. Node embeddings for the new nodes need to be computed rapidly as well. Using the entire graph structure in a transductive manner to compute the embeddings may not scale well. Researchers (Rath, et al., 2020) therefore used an inductive approach

inspired by the GraphSAGE work (Hamilton et al., 2017) to compute embeddings based on the inductive representation learning. The graph models how fake news spreads on microblog sites like Twitter.

Reinforcement Learning

A predominant assumption in most machine learning models is that the data is i.i.d, Independent and Identically Distributed. Most of the machine learning solutions presented in the literature for the misinformation containment problem focus on a dataset from a single domain such as politics (Pendyala et al., 2018) or healthcare (Pendyala & Figueira, 2015), the embeddings for which are i.i.d. However, misinformation often covers multiple domains. Reinforcement learning can be used for cross-domain modeling (Mosallanezhad et al., 2022). Users' comments, interactions, and information in two disparate domains are used to learn a domain-agnostic representation of the information to aid in its classification. Reinforcement learning is used to convert the representations in the source domain into a representation in the target domain. Reinforcement learning has also been used to increase the availability of labeled data (Wang, et al., 2020). Annotating information is an expensive process and requires manual expertise. An automatic annotator assigns the labels based on user reports. This initial labeling is weak and not entirely accurate. Reinforcement learning is then used to select the best-labeled instances from the weakly labeled data.

Semi-Supervised Learning

Similar to the above reinforcement learning approach to augment the labeled data, researchers (Li, et al., 2022) introduced a "confidence network layer" into a bidirectional LSTM to filter out the data that is confidently labeled. The neural network is initially trained on a limited labeled dataset in a supervised manner. The confidence network layer adds the confidently labeled data to this initial dataset in an iterative manner, quite along the lines of semi-supervised learning. Semi-supervised learning is often used in conjunction with graphs. Typically, each data item is a node, and the weighted edges connecting them indicate the similarity between the nodes. Labels can then propagate from labeled nodes to unlabeled ones. A graphs-based semi-supervised learning approach has also been used for fake news detection (Benamira, 2019). A graph is constructed from the GloVe embeddings of the documents. Each node is a document. Some of the nodes are already labeled as genuine, some others are labeled fake, and several others are unlabeled. The nodes are interconnected based on the k-nearest neighbors algorithm. Graph convolution network approach is used to classify the unlabeled nodes.

Transfer Learning

Pretrained language models like BERT, RoBERTa, GPT2, and Funnel generate embeddings that can be used subsequently as inputs to Artificial Neural Networks (ANN) or Convolutional Neural Networks (CNN) to determine the veracity of a given text (Samadi, Mousavian, & Momtazi, 2021). Capsule networks can also be used with word embeddings in the process of transfer learning (Goldani, Momtazi, & Safabakhsh, 2021). The datasets used for the problem are well-known and limited in scope. The literature survey confirms that there have not been any attempts to generate embeddings exclusively from misinformation.

DETECTING BOTS

Misinformation spread often happens on social media using accounts owned by software. The trend has given rise to the growth of companies that even offer Bot-as-a-service (BAAS). During the 2016 US Presidential elections, about 20% of the social media discussions on the topic happened using bots (Wu et al., 2018). A wide spectrum of various types of machine learning algorithms can be used to detect bots. Algorithms vary, but the overall process of classification of the user as a bot or not is somewhat similar when the machine learning is supervised. Various features such as the screen name, description, and the number of followers are extracted from the user profiles. Parametric, supervised machine learning algorithms such as logistic regression determine the weights for each of these features from already classified data. The algorithms model the classification as a function of the weighted features. The model can then be applied to new user data, based on which the user needs to be classified as a bot or human.

Obtaining user profile data that is already classified as belonging to a bot or a human being may pose challenges. In the absence of such a training set, unsupervised learning algorithms can be used to cluster the data into those belonging to bots and others to real humans. Researchers (Wu et al., 2018) use several clustering algorithms on network traffic to detect if a host is infected by bots. Similar techniques can be used on social media profiles as well. Hegelich and Janetzko (Hegelich & Janetzko, 2016) apply k-means and hierarchical clustering algorithms to the posts made by social bots in Ukraine elections to draw interesting conclusions about the behavioral patterns of the social bots. Clustering algorithms still work with the features extracted from the data and group the data into clusters based on similarity.

For certain classification problems, deep learning can achieve better classification accuracy, as noted in the preceding section on fake profile identification. Authors (Kudugunta & Ferrara, 2018) have proven that the problem of classifying accounts as bots or humans can be solved with superlative accuracies using deep neural networks, but this time, using a different architecture than CNNs. Using Long Short-Term Memory (LSTM) architecture, the authors bring out the advantage of using deep learning techniques over conventional machine learning algorithms. The LSTM architecture can be applied even when the feature set is small and the size of the dataset is limited, as the authors show. LSTM networks are a type of Recurrent Neural network (RNN), which provide for information to persist. RNNs have been extensively used to come up with outstanding solutions to challenging problems.

Generative Adversarial Networks

Machine learning and deep learning particularly have been central to some of the techniques discussed earlier. One area of deep learning that requires special mention in the context of trusting social media is Generative Adversarial Network or GAN. GAN models have helped create unbelievably realistic fake content that includes images and text. The website, ThisPersonDoesNotExist.com displays several faces generated using GANs that look amazingly realistic but are completely synthesized. The fake facial images created using GAN models are used on social media to create fake accounts to propagate malicious agendas. Since the fake content generated by GAN is close to real, it becomes a challenge to differentiate the fake ones from the real images. GAN can be used for addressing the problem of trust in social media in two ways: (a) to generate a dataset of fake content for training the model used to detect new fake content and (b) to detect fake content itself.

Multiple techniques have been developed to deal with the problem of detecting fakes generated by GAN and one of them is using part of GAN itself. A GAN has two parts – a generator and a discriminator, both implemented as neural networks, typically convolutional or recurrent. The generator creates fake content and the discriminator keeps rejecting it until it is convinced that the content from the generator is real. The function of the discriminator is really to reject fake content. Hence, GAN discriminators can potentially be used to detect fake content (Marra et al., 2018). The authors use the same GAN discriminator that was used in the process of generating fake images to train various other models described in the paper, as a baseline algorithm to detect fake content. To make sure that the discriminator is not biased because it has already gone through the training samples, the discriminator needs to be retrained. The work uses several other models as well and compares the results with those obtained by using the GAN discriminator.

Given that majority of the social media is still in text format, a major development in detecting misleading content is to apply the powerful GAN models to text data (Aghakhani et al., 2018). Here too, the discriminator part of the GAN does the job of determining fake content, this time, textual reviews on TripAdvisor with an accuracy of 89.1%. Since Convolutional Neural Network (CNN) works better with text, the discriminator is implemented as a CNN. The authors use two discriminators, both CNN, instead of just one that the GAN model originally proposed. Based on the experiments and results obtained from them, they confirm that using two discriminators works better in this case of textual reviews.

Interpretable models, which can explain the classification done by the model are the need of the hour. Explanations provide transparency and better confidence in the model. Authors (Carton et al., 2018) use the GAN philosophy to develop "Extractive Adversarial Networks" to go a step further, beyond text classification, to provide explanations for the classification decisions. Their work detects comments on social media that are personal attacks and points out the words in those comments that are the reason the model classified the comment as a personal attack. The key difference between their model from the original GAN is that their model extracts a modified sample from an existing sample instead of generating one, as done in GAN. The generation function is thus replaced by extraction, hence the name they chose for the model.

MISINFORMATION CONTAINMENT IS STILL UNSOLVED: WHY?

As the previous sections indicate, substantial research and success have been reported in the literature so far. However, there is no ambiguity in stating that the problem is largely unsolved. To understand the reasons, there is a need to understand the nature of misinformation. Unlike spam, which can be detected based on the occurrence of certain word patterns, misinformation is highly complex. It is hugely a challenge even for human beings to detect misinformation. For instance, it can take years for the truth to be established in a court of law. The following paragraphs discuss a few points to consider when designing solutions to misinformation containment.

Truth is Temporal, Subjective, and Relative

There were times when the truth about the earth was that it is flat, people who believe that the truth is that God does not exist, and objects that are small only when compared to other relatively large objects.

On the other hand, machine learning models used in the current literature are fixed and cannot handle the dynamic variations in subjectivity or relativity.

Determining Truth can Require a huge Corpus of Prior Knowledge

Often, the truth can be determined only after cross-checking against a huge corpus of facts, evidence, and reasoning. Machine learning models are incapable of such cross-checking as compared to First-Order-Logic (FOL) and other formal methods that have the implements to reason and inference from prior statements.

Truth can evade Feature Engineering

The same source of information with their features intact can produce conflicting statements. For instance, there are several cases where a user on Twitter posted conflicting tweets. Merely extracting the features, whether manually engineered or automatically generated by the hidden layers via deep learning are unlikely to flag the misinformation. The feature set used in a substantial part of the literature is limited to temporal or contextual or content-based, whereas the need is for much more comprehensive and exhaustive information.

Ingenuity in Camouflaging Misinformation

Misinformation is often seamlessly interwoven with truth in ingenious ways. Even the intent and purport can sound genuine. Latent space mapping and self-supervised learning approaches can help only to some extent but are not always accurate or exhaustive.

Limitations of the Machine Learning Models

RNNs, CNNs, and other language models that are often used for Natural Language Understanding (NLU) in the process of misinformation detection cannot capture long-term dependencies in large texts. Even the latest transformer-based models like BERT are only good for 512 tokens in the text and efforts were made to increase it to 2048 tokens (Yang, et. al., 2020).

There is no Silver Bullet

Misinformation containment is not a single problem to have a single do-it-all solution. Misinformation is manifested in many modalities and forms. Each manifestation needs to be addressed in one or more ways. The current approaches address the problem by experimenting with a sample dataset, which is usually identically distributed, report the results from the same dataset and portray general success based on those results. On the other hand, misinformation is far from being identically distributed.

FUTURE RESEARCH DIRECTIONS

Despite its limitations and extensive use, it does not appear that machine learning has been fully used in addressing the misinformation containment problem. For instance, unsupervised and semi-supervised techniques do not seem to have been used in areas they can be used, such as pointed out in the subsection on "Detecting Bots". Given that most of the available data in the social media space cannot easily be classified with certainty into bot-generated or genuinely generated by humans, unsupervised and semi-supervised methods should show substantial promise. State-of-the-art deep learning frameworks in general are computationally expensive and require humongous data. Once trained, the essential parameters are fixed, so newer patterns are not modeled if the data is still evolving. Static feature extraction for subsequent classification using parametric methods does not model the evolving nature of misinformation either. On the other hand, Misinformation Containment (MC) needs to be continuous, instantaneous, evolve with the changing patterns in the data, and use fewer resources so that users can deploy the models on edge devices such as smartphones. Future research needs to address these characteristics.

The research needs to focus on both applications of the existing machine learning frameworks and changing the underlying methodologies to suit the needs of misinformation containment. Graph theoretic approaches hold substantial promise because they can capture dependencies in a long sequence substantially well and can serve as the long-term memory that RNNs fail to provide. Inductive approaches to compute embeddings such as GraphSAGE (Hamilton et al., 2017) can scale well. The problem indeed needs to be solved in multiple stages as pointed out in the Fake News Challenge -1 (FNC-1)1 problem description.

CONCLUSION

Misinformation containment is a complex problem. The problem has been addressed using several techniques, algorithms, and frameworks available in machine learning, but the problem remains unsolved. Similar problems such as spam, viruses, cyberattacks, and other malicious implements are in reasonable control, but not misinformation, which is a reflection of the complexity of the problem. This chapter presented a brief survey of the trends in misinformation containment using machine learning and explained why the current approaches have not been effective. Future research directions should go beyond the traditional dataset collection, train, validate, and test cycle, and use frameworks that can model the characteristics of misinformation better.

ACKNOWLEDGMENT

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

Aghakhani, H., Machiry, A., Nilizadeh, S., Krügel, C., & Vigna, G. (2018). Detecting Deceptive Reviews Using Generative Adversarial Networks. *2018 IEEE Security and Privacy Workshops (SPW)*, 89-95.

Antony Vijay, J., Anwar Basha, H., & Arun Nehru, J. (2021). A Dynamic Approach For Detecting The Fake News Using Random Forest Classifier And Nlp. In *Computational Methods And Data Engineering* (pp. 331–341). Springer. doi:10.1007/978-981-15-7907-3_25

Benamira, A., Devillers, B., Lesot, E., Ray, A. K., Saadi, M., & Malliaros, F. D. (2019, August). Semi-supervised learning and graph neural networks for fake news detection. In *2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 568-569). IEEE. 10.1145/3341161.3342958

Borges, L., Martins, B., & Calado, P. (2019). Combining similarity features and deep representation learning for stance detection in the context of checking fake news. *Journal of Data and Information Quality*, *11*(3), 1–26. doi:10.1145/3287763

Braşoveanu, A. M., & Andonie, R. (2019). *Semantic fake news detection: a machine learning perspective*. International Work-Conference on Artificial Neural Networks.

Budak, C., Agrawal, D., & El Abbadi, A. (2011). Limiting The Spread Of Misinformation In Social Networks. *Proceedings Of The 20th International Conference On World Wide Web*, 665–674. 10.1145/1963405.1963499

Carton, S., Mei, Q., & Resnick, P. (2018). *Extractive Adversarial Networks: High-Recall Explanations For Identifying Personal Attacks In Social Media Posts*. doi:10.18653/v1/D18-1386

Deepak, S., & Chitturi, B. (2020). Deep neural approach to Fake-News identification. *Procedia Computer Science*, *167*, 2236–2243. doi:10.1016/j.procs.2020.03.276

Elsherief, M., Sumner, S. A., Jones, C. M., Law, R. K., Kacha-Ochana, A., Shieber, L., Cordier, L., Holton, K., & De Choudhury, M. (2021). Characterizing And Identifying The Prevalence Of Web-Based Misinformation Relating To Medication For Opioid Use Disorder: Machine Learning Approach. *Journal of Medical Internet Research*, *23*(12), E30753. doi:10.2196/30753 PMID:34941555

Ghafari, S. M., Joshi, A., Beheshti, A., Paris, C., Yakhchi, S., & Orgun, M. (2019). Dcat: A Deep Context-Aware Trust Prediction Approach For Online Social Networks. *Proceedings Of The 17th International Conference On Advances In Mobile Computing & Multimedia*, 20–27. 10.1145/3365921.3365940

Goldani, M. H., Momtazi, S., & Safabakhsh, R. (2021). Detecting fake news with capsule neural networks. *Applied Soft Computing*, *101*, 106991. doi:10.1016/j.asoc.2020.106991

Ghayoomi, M., & Mousavian, M. (2022). Deep Transfer Learning For Covid-19 Fake News Detection In Persian. *Expert Systems: International Journal of Knowledge Engineering and Neural Networks*, *39*(8), E13008. doi:10.1111/exsy.13008 PMID:35599852

Guderlei, M., & Aßenmacher, M. (2020, December). Evaluating unsupervised representation learning for detecting stances of fake news. In *Proceedings of the 28th International Conference on Computational Linguistics* (pp. 6339-6349). 10.18653/v1/2020.coling-main.558

Hakak, S., Alazab, M., Khan, S., Gadekallu, T. R., Maddikunta, P. K., & Khan, W. Z. (2021). An ensemble machine learning approach through effective feature extraction to classify fake news. *Future Generation Computer Systems*, *117*, 47–58. doi:10.1016/j.future.2020.11.022

Hamilton, W., Ying, Z., & Leskovec, J. (2017). Inductive Representation Learning On Large Graphs. *Advances in Neural Information Processing Systems*, 30.

Hasani, R., Lechner, M., Amini, A., Rus, D., & Grosu, R. (2021). Liquid time-constant networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(9), 7657–7666. doi:10.1609/aaai.v35i9.16936

He, Q., Lv, Y., Wang, X., Huang, M., & Cai, Y. (2022). Reinforcement Learning-Based Rumor Blocking Approach In Directed Social Networks. *IEEE Systems Journal*, 1–11. doi:10.1109/JSYST.2022.3159840

Hegelich, S., & Janetzko, D. (2016). Are Social Bots On Twitter Political Actors? Empirical Evidence From A Ukrainian Social Botnet. *Tenth International Aaai Conference On Web And Social Media*.

Huh, M., Liu, A., Owens, A., & Efros, A. A. (2018). Fighting Fake News: Image Splice Detection Via Learned Self-Consistency. *Proceedings Of The European Conference On Computer Vision (Eccv)*, 101–117. 10.1007/978-3-030-01252-6_7

Karpov, I., & Glazkova, E. (2020). Detecting Automatically Managed Accounts In Online Social Networks: Graph Embeddings Approach. *International Conference On Analysis Of Images, Social Networks And Texts*, 11–21.

Kozik, S., Kula, S., Choraś, M., & Woźniak, M. (2022). Technical Solution To Counter Potential Crime: Text Analysis To Detect Fake News And Disinformation. *Journal of Computational Science*, *60*, 101576. doi:10.1016/j.jocs.2022.101576

Kudugunta, S., & Ferrara, E. (2018). Deep Neural Networks For Bot Detection. *Information Sciences*, *467*, 312–322. doi:10.1016/j.ins.2018.08.019

Li, X., Lu, P., Hu, L., Wang, X., & Lu, L. (2022). A Novel Self-Learning Semi-Supervised Deep Learning Network To Detect Fake News On Social Media. *Multimedia Tools and Applications*, *81*(14), 19341–19349. doi:10.100711042-021-11065-x PMID:34093070

Lo, K.-C., Dai, S.-C., Xiong, A., Jiang, J., & Ku, L.-W. (2022). Victor: An Implicit Approach To Mitigate Misinformation Via Continuous Verification Reading. *Proceedings Of The Acm Web Conference 2022*, 3511–3519. 10.1145/3485447.3512246

Lwowski, B., & Najafirad, P. (2020). *Covid-19 surveillance through twitter using self-supervised and few shot learning*. Academic Press.

Marra, F., Gragnaniello, D., Cozzolino, D., & Verdoliva, L. (2018). Detection Of Gan-Generated Fake Images Over Social Networks. *2018 IEEE Conference On Multimedia Information Processing And Retrieval (Mipr),* 384–389.

Mondal, S. K., Sahoo, J. P., Wang, J., Mondal, K., & Rahman, M. (2022). Fake News Detection Exploiting Tf-Idf Vectorization With Ensemble Learning Models. In *Advances In Distributed Computing And Machine Learning* (pp. 261–270). Springer.

Mosallanezhad, A., Karami, M., Shu, K., Mancenido, M. V., & Liu, H. (2022). Domain Adaptive Fake News Detection Via Reinforcement Learning. *Proceedings Of The Acm Web Conference 2022*, 3632–3640. 10.1145/3485447.3512258

O'Brien, K., Simek, O., & Waugh, F. (2019). *Collective Classification For Social Media Credibility Estimation*. Academic Press.

Paka, W. S., Bansal, R., Kaushik, A., Sengupta, S., & Chakraborty, T. (2021). Cross-SEAN: A cross-stitch semi-supervised neural attention model for COVID-19 fake news detection. *Applied Soft Computing*, *107*, 107393. doi:10.1016/j.asoc.2021.107393

Palani, B., Elango, S., & Viswanathan, K. (2022). Cb-Fake: A Multimodal Deep Learning Framework For Automatic Fake News Detection Using Capsule Neural Network And Bert. *Multimedia Tools and Applications*, *81*(4), 5587–5620. doi:10.100711042-021-11782-3 PMID:34975284

Patel, A., & Meehan, K. (2021). Fake News Detection On Reddit Utilising Countvectorizer And Term Frequency-Inverse Document Frequency With Logistic Regression, Multinominalnb And Support Vector Machine. *2021 32nd Irish Signals And Systems Conference (Issc)*, 1–6.

Pendyala, V. S., & Figueira, S. (2015, October). Towards a truthful world wide web from a humanitarian perspective. In *2015 IEEE Global Humanitarian Technology Conference (GHTC)* (pp. 137-143). IEEE. 10.1109/GHTC.2015.7343966

Pendyala, V. (2018). *Veracity of Big Data: Machine Learning and Other Approaches to Verifying Truthfulness*. Apress. doi:10.1007/978-1-4842-3633-8

Pendyala, V. S., Liu, Y., & Figueira, S. M. (2018). A Framework For Detecting Injected Influence Attacks On Microblog Websites Using Change Detection Techniques. *Development Engineering*, *3*, 218–233. doi:10.1016/j.deveng.2018.08.002

Pendyala, V. S. (2019). Securing Trust In Online Social Networks. *International Conference On Secure Knowledge Management In Artificial Intelligence Era*, 194–201.

Rajalaxmi, R., Narasimha Prasad, L., Janakiramaiah, B., Pavankumar, C., Neelima, N., & Sathishkumar, V. (2022). Optimizing Hyperparameters And Performance Analysis Of Lstm Model. In *Detecting Fake News On Social Media*. Transactions On Asian And Low-Resource Language Information Processing.

Rath, B., Salecha, A., & Srivastava, J. (2020, December). Detecting fake news spreaders in social networks using inductive representation learning. In *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 182-189). IEEE. 10.1109/ASONAM49781.2020.9381466

Sahan, M., Smidl, V., & Marik, R. (2021). Active Learning For Text Classification And Fake News Detection. *2021 International Symposium On Computer Science And Intelligent Controls (Iscsic)*, 87–94. 10.1109/ISCSIC54682.2021.00027

Salem, F. K., Al Feel, R., Elbassuoni, S., Ghannam, H., Jaber, M., & Farah, M. (2021). Meta-learning for fake news detection surrounding the Syrian war. *Patterns*, *2*(11), 100369. doi:10.1016/j.patter.2021.100369 PMID:34820650

Samadi, M., Mousavian, M., & Momtazi, S. (2021). Deep contextualized text representation and learning for fake news detection. *Information Processing & Management*, *58*(6), 102723. doi:10.1016/j.ipm.2021.102723

Shen. (2022). Mdn: Meta-Transfer Learning Method For Fake News Detection. *Ccf Conference On Computer Supported Cooperative Work And Social Computing*, 228–237.

Urena, R., Chiclana, F., & Herrera-Viedma, E. (2020). Decitrustnet: A Graph Based Trust And Reputation Framework For Social Networks. *Information Fusion*, *61*, 101–112. doi:10.1016/j.inffus.2020.03.006

Wanda, P., & Jie, H. J. (2020). Deepprofile: Finding Fake Profile In Online Social Network Using Dynamic Cnn. *Journal Of Information Security And Applications*, *52*, 102465. doi:10.1016/j.jisa.2020.102465

Wang, Y., Yang, W., Ma, F., Xu, J., Zhong, B., Deng, Q., & Gao, J. (2020, April). Weak supervision for fake news detection via reinforcement learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(01), 516–523. doi:10.1609/aaai.v34i01.5389

Wu, W., Alvarez, J., Liu, C., & Sun, H.-M. (2018). Bot Detection Using Unsupervised Machine Learning. *Microsystem Technologies*, *24*(1), 209–217. doi:10.100700542-016-3237-0

Xie, J., Chai, Y., & Liu, X. (2022). An Interpretable Deep Learning Approach To Understand Health Misinformation Transmission On Youtube. *Proceedings Of The 55th Hawaii International Conference On System Sciences*. 10.24251/HICSS.2022.183

Yang, L., Zhang, M., Li, C., Bendersky, M., & Najork, M. (2020, October). Beyond 512 tokens: Siamese multi-depth transformer-based hierarchical encoder for long-form document matching. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management* (pp. 1725-1734). 10.1145/3340531.3411908

ENDNOTE

[1] http://www.fakenewschallenge.org