Université de Montréal

Personalized Fake News Aware Recommendation System

par

Dorsaf Sallami

Département d'informatique et de recherche opérationnelle
Faculté des arts et des sciences

Mémoire présenté en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en Informatique

August 12, 2022

Université de Montréal

Faculté des arts et des sciences

Ce mémoire intitulé

Personalized Fake News Aware Recommendation System

présenté par

Dorsaf Sallami

a été évalué par un jury composé des personnes suivantes :

Claude Frasson
(président-rapporteur)

Esma Aimeur
(directeur de recherche)

Jian-Yun Nie
(membre du jury)

Résumé

De nos jours, où les actualités en ligne sont si répandues, diverses méthodes ont été développées afin de fournir aux utilisateurs des recommandations d'actualités personnalisées. De merveilleuses réalisations ont été faites lorsqu'il s'agit de fournir aux lecteurs tout ce qui pourrait attirer leur attention. Bien que la précision soit essentielle dans la recommandation d'actualités, d'autres facteurs, tels que la diversité, la nouveauté et la fiabilité, sont essentiels pour satisfaire la satisfaction des lecteurs. En fait, les progrès technologiques apportent des défis supplémentaires qui pourraient avoir un impact négatif sur le domaine de l'information. Par conséquent, les chercheurs doivent tenir compte des nouvelles menaces lors de l'élaboration de nouvelles recommandations. Les fausses nouvelles, en particulier, sont un sujet brûlant dans les médias aujourd'hui et une nouvelle menace pour la sécurité publique.

Au vu des faits mentionnés ci-dessus, ce travail présente un système modulaire capable de détecter les fausses nouvelles, de recommander des nouvelles à l'utilisateur et de les aider à être plus conscients de ce problème. Tout d'abord, nous suggérons FANAR, FAke News Aware Recommender system, une modification d'algorithme de recommandation d'actualités qui élimine les personnes non fiables du voisinage de l'utilisateur candidat. A cette fin, nous avons créé un modèle probabiliste, Beta Trust Model, pour calculer la réputation des utilisateurs. Pour le processus de recommandation, nous avons utilisé Graph Neural Networks. Ensuite, nous proposons EXMULF, EXplainable MUltimodal Content-based Fake News Detection System. Il s'agit de l'analyse de la véracité de l'information basée sur son contenu textuel et l'image associée, ainsi qu'un assistant d'intelligence artificielle Explicable (XAI) pour lutter contre la diffusion de fake news. Enfin, nous essayons de sensibiliser aux fake news en fournissant des alertes personnalisées basées sur le profil des utilisateurs.

Pour remplir l'objectif de ce travail, nous construisons un nouveau jeu de données nommé FNEWR. Nos résultats expérimentaux montrent qu'EXMULF surpasse 10 modèles de pointe de détection de fausses nouvelles en termes de précision. Aussi, FANAR qui prend en compte les informations visuelles dans les actualités, surpasse les approches concurrentes basées uniquement sur le contenu textuel. De plus, il permet de réduire le nombre de fausses nouvelles dans la liste des recommandations.

Mots clés : Recommandations d'actualités personnalisées, Détection de fake news, Données multimodales, IA explicable, réputation des utilisateurs, sensibilisation

Abstract

In today's world, where online news is so widespread, various methods have been developed in order to provide users with personalized news recommendations. Wonderful accomplishments have been made when it comes to providing readers with everything that could attract their attention. While accuracy is critical in news recommendation, other factors, such as diversity, novelty, and reliability, are essential in satisfying the readers' satisfaction. In fact, technological advancements bring additional challenges which might have a detrimental impact on the news domain. Therefore, researchers need to consider the new threats in the development of news recommendations. Fake news, in particular, is a hot topic in the media today and a new threat to public safety.

This work presents a modularized system capable of recommending news to the user and detecting fake news, all while helping users become more aware of this issue. First, we suggest FANAR, FAke News Aware Recommender system, a modification to news recommendation algorithms that removes untrustworthy persons from the candidate user's neighbourhood. To do this, we created a probabilistic model, the Beta Trust model, to calculate user reputation. For the recommendation process, we employed Graph Neural Networks. Then, we propose EXMULF, EXplainable MUltimodal Content-based Fake News Detection System. It is tasked with the veracity analysis of information based on its textual content and the associated image, together with an Explainable AI (XAI) assistant that is tasked with combating the spread of fake news. Finally, we try to raise awareness about fake news by providing personalized alerts based on user reliability.

To fulfill the objective of this work, we build a new dataset named FNEWR. Our experiments reveal that EXMULF outperforms 10 state-of-the-art fake news detection models in terms of accuracy. It is also worth mentioning that FANAR , which takes into account visual information in news, outperforms competing approaches based only on textual content. Furthermore, it reduces the amount of fake news found in the recommendations list.

Keywords: Personalized news recommendations, Fake news detection, Multimodal data, Explainable AI, User Reputation, Awareness

Contents

List of tables

List of figures

List of Abbreviations and Acronyms

AGNN	Attention-based Graph Neural Network
AMRAN	Multi-Relational Attention Network
AT	Adaptive Tag algorithm
BDANN	BERT-based Domain Adaptation Neural Network
BERT	Bidirectional Encoder Representations from Transformers
CBF	Content-Based Filtering
CF	Collaborative Filtering
CNN	Convolutional Neural Networks
DAGA-NN	Domain-Adversarial and Graph-Attention Neural Network
EXMULF	Explainable Multimodal Content-based Fake News Detection System

FANAR	FAke News Aware Recommender system
FNASY	Fake News Awareness System
FNEWR	Fake NEWs Recommendation
GERL	Graph Enhanced Representation Learning
GNewsRec	Graph Neural News Recommendation
GNN	Graph Neural Networks
GNUD	Graph Neural News Recommendation Model with Unsupervised Preference Disentanglement
HCS	Hybrid Crowd Signals
HMCAN	Hierarchical Multimodal Contextual Attention Network
IGNN	Interaction Graph Neural Network
KCNN	Knowledge-aware Convolutional Neural Networks
KNN	K-Nearest Neighbors

LDA	Latent Dirichlet Allocation
LIME	Local Interpretable Model-agnostic Explanations
LSTM	Long Short-Term Memory
LXMERT	Learning Cross-Modality Encoder Representations from Transformers
MCNN	Multimodal Consistency Neural Network
MMR	Maximal Marginal Relevance
MM-Rec	MultiModal news RECommendation system
MRNT	model for news tags reconstructing
MVL	Multi-View Learning
NLP	Natural Language Processing
NRS	News Recommendation System
PCNN	Pulse Coupled Neural Network

RA	Recommendation Algorithm
ResNet	Residual neural Network
RG	Research Goal
RL	Reinforcement Learning
ROIs	Regions Of Interest
RS	Recommendation System
SAFE	Similarity-Aware Multi-Modal Fake News Detection
TD	Topic diversification
TSHGNN	Time-Sensitive Heterogeneous Graph Neural Network
VGG	Visual Geometry Group
XAI	Explainable Artificial Intelligence

Acknowledgments

I would like to extend my special thanks and sincere gratitude to all the people who have contributed to this endeavour.

I would like to start by thanking my professor and advisor Esma Aïmeur for the constant follow-ups all along the way, the excellent scientific guidance and the constructive criticisms that have greatly enriched my knowledge, sharpened my view, showed me the right path, and provided me with invaluable assistance.

I am grateful to all my professors who sculpted the knowledge of my peers and me, and have put huge efforts into our education. Without them, I would not have been able to conduct this work.

I would also like to express my immeasurable gratitude to all the members of the jury for reviewing and evaluating this work.

Last, but not least, my family deserves endless gratitude. To my family, I owe everything, including this.

Part I

General Context

Chapter 1

Introduction

Introduction

In this chapter, we provide a general overview of the context in which the project resides. We start by exploring the motivations. The thesis statement and research objectives are then presented. Finally, we go through the report structure.

1.1. Motivation

Under the surface of online technologies exists a deep sea of algorithms. They impact how we consume data and how information or services are recommended to us, all while hiding in plain sight.

In today's world, where recommendations are so widespread, it's easy to ignore the seeming simplicity with which these algorithms, created to enhance our consumer choices, have been integrated into almost every device and platform. With the increasing spread of online information and services, customers are finding it challenging to make clear decisions on a variety of products such as news, movies, music, and books. The goal of recommendation systems is to meet the specific user preferences by presenting specialized items, thus solving the issue of information overload [**40**].

News recommendation systems (NRS), particularly, have received a lot of interest. Because of its convenience and recency, more and more individuals choose to read news online rather than in paper format such as printed press. However, a massive number of news events may be released at a rate of hundreds, if not thousands, every hour. A difficult task lies in determining how to choose efficiently between specific news items to recommend to readers from a massive corpus of newly published press releases, where the recommended news items should fit the reader's reading preference as accurately as possible. This issue refers to personalized news recommendation. Recently, personalized news recommendation has become a promising research path as the Internet provides fast access to real-time information from multiple sources around the world. Existing personalized news recommendation

systems strive to adapt their services to individual users by virtue of both user and news content information.

A recommender system's purpose is to anticipate how likely users are to appreciate unknown items based on the data the system has about them. However, evaluating RS based only on accuracy cannot provide a response to the issue of whether users are satisfied with the recommendations or not. This type of challenge requires the consideration of other aims for a recommender system, which might address factors beyond just accuracy. While accuracy is essential, the quality of news recommendations cannot be enhanced until non-accuracy factors are considered. Other factors, such as innovation, diversity, unexpectedness, and coverage, are just as significant as accuracy for user satisfaction [**68**]. Therefore, one additional challenge, besides accurately predicting whether an article is relevant for a user or not, is to take additional qualitative factors into account.

It is a wonderful accomplishment to provide readers with content that catches their attention. However, relying only on machine learning algorithms, as in recommender systems, is fraught with risk. On the one side, researchers are attempting to enhance the news ecosystem through the development of various algorithms and solutions. On the other side, technology is introducing new challenges, which we might refer to as 'post algorithmic' news issues. These difficulties are thought to have a detrimental impact on the development of news recommendations (fake news, exaggerated news, racism, persecution, stereotypes, and so on), users' psychological behaviour, consumption habits, and the overall user experience with NRS [**68**].

The world wide web has introduced new threats to public safety and the well-being of society as a whole. Fake news, notably, is a hot topic in the media today. It is easier than ever to broadcast disinformation, see it spread, and then watch it bring down companies, destroy reputations, and destabilize political figures. The issues of online misinformation and fake news have grown in significance in an era where user-generated content and social media platforms are powerful influences in creating and spreading news stories. Untrustworthy information and deceptive content are often uploaded and widely spread on prominent social media platforms. This phenomenon is a real problem that requires real actions in order to be eliminated. Generally, researchers work on fake news detection in order to categorise information as true or false [**98, 105, 98, 108**]. However, it is clear that the problem necessitates far more effort than just detecting it. Identifying fake news is a critical step in avoiding the problem, but the situation warrants further education and awareness endeavors, to assist people in preventing fake news in their daily lives.

Finally, a NRS could be presented as a tool to aid individuals in such a dilemma, with the goal of creating a system that allows people to acquire balanced news information. As a result, in addition to effectively anticipating whether or not an article is suitable for a user, another problem to overcome is the inclusion of quality-oriented features, such as assisting customers in spotting fake news.

1.2. Thesis Statement

In this thesis, we aim to create a system that provides personalized news recommendations while assisting users in the avoidance of fake news and its dissemination, as well as raising awareness about the issue. As a result, the system includes recommendation, fake news detection, and awareness components.

However, misinformation spreads online due to a variety of factors, including how information is transmitted, the digital platforms where it is dispersed, and the algorithms that manage information suggestions within those platforms. Also, recommendation algorithms are responsible for the propagation of disinformation. Since recommendation algorithms have been severely criticized for being unintentional ways of amplifying and spreading misinformation [**27**], researchers explore the effect of existing algorithms on the recommendation of inaccurate and misleading content. Furthermore, automatically detecting fake news is a difficult process. First, humans themselves are naturally poor at distinguishing between true and false news [**77**], especially when it comes to sensitive topics like politics and health. Second, news articles are generated by several sources, each with their own content style and inherent biases, and they are transmitted in various ways through separate environments, making the process of identifying fake news even more difficult.

Therefore, we aim to investigate possible modifications and solutions to the recommendations process that may be beneficial when considering the issues posed by fake news.

1.3. Research Goals

The overall purpose of this thesis can be divided into the specific research goals (RGs) listed below.

- **RG1 - Adapting recommendation algorithm to avoid fake news:**
  The recommendation systems play a crucial role in information dissemination and propagation [**27**]. This is especially true for large-scale platforms like social media, where recommender systems help users get access to vast amounts of user-generated information. Furthermore, the majority of approaches in scientific literature seek to discover misleading material that has already been propagated on social media. Moreover, learning accurate news representations is the backbone of news recommendation [**93**]. The majority of current news representation systems derive news representations only from news texts [**5, 38, 86**], while ignoring the visual information. However, when posted on news websites, many news articles are accompanied with images in addition to their text content. In fact, visual information (video, image) provides additional information to better understand news content and enhance news recommendations.

We address all these issues in this research step. Specifically, we propose to adapt recommendation algorithms in a way that recommendations are solely based on trustworthy users. To accomplish this task, we provide a probabilistic model to calculate user reputation based on explicit user opinion. Furthermore, we develop a model that incorporates both text and image pieces of information into news representation learning and illustrates their relatedness for better news content understanding, when image-related information is typically neglected in existing news recommendation approaches. Additionally, our method can simulate the cross-modal relationship between text and image. As a result, our algorithm performs better in terms of news recommendation.

- **RG2 - Exploring the multimodal data available in news content to detect fake news and provide explanations:**
  Online articles and posts often include images that usually attract the attention of the users. The images may be used to help a classification system. Furthermore, the similarity between the image and the text is very important since it is possible that in some fake news the image to be contradictory to the content. Also, it is possible that images in fake and real news follow different patterns or that have been modified in order to attract users' attention [**31**]. As a consequence, investigating multimodal data may enhance fake news detection results.
  Numerous attempts have been made to build deep learning-based automatic fake news detection methods. However, there has been little previous work that has moved beyond the black-box characteristic of such systems and focused on offering explanations to users of online social networks. Such explanations are critical to reflect news credibility, increase users' knowledge, and potentially influence their behaviour when it comes to preserving both individuals' and society's security and privacy. Exploring the multimodal data available in news materials, on the other hand, is critical for enhancing the explanations supplied to users and for detecting fake news. Many real-world applications do not rely on a single data modality. Text, images, videos, and other media can be found on websites, for example. If we limit ourselves to using only one modality, we will lose all of the knowledge contained in the others.
  According to studies, it is still difficult for humans to judge the reliability of a specific piece of news information based purely on automatic models and without further explanations [**63, 46**]. Additionally, humans achieved an average accuracy of 54% in the testing of deception judgment [**13**]. As a result, in recent years, identifying fake news has migrated to explainable and interpretable automated detection methods.

Thus, we intend to develop a content-based fake news detection module, explore multimodal data and give meaningful explanations, and integrate explainable artificial intelligence (XAI).

- **RG3 - Raising awareness about fake news:**
  The great majority of awareness studies emphasize the need to be aware of misleading information rather than providing strategies for increasing individual awareness. To limit the spread of fake news, people must first be notified of it. Warning them that the news is misleading and providing appropriate explanations may persuade them not to read or spread it. Hence, aspects of awareness should be efficiently incorporated into the recommendation process.
  In our case, we include an awareness module in the post-recommendation part. Personalized awareness is provided based on the user's profile.

1.4. Contributions

The main contributions of this thesis are summarized as follows:

- **Research goal 1**:
  In order to achieve the first study objective, we modify recommendation algorithm such that recommendations are entirely based on trustworthy people. We have the following contributions for this purpose:

  (i) A new dataset. We perform a comprehensive assessment of the available datasets. However, these did not meet our requirements for the development of the system. As a result, we built our own dataset for the purpose of this study and as a first step forward.

  (ii) We provide a new probabilistic model for user reputation in the context of fake news.

  (iii) We provide a novel collaborative filtering strategy for the news recommendation assignment that may considerably prevent the propagation of fake news by avoiding untrustworthy neighbors.

  (v) We propose a multimodal way of representing news. The majority of the studies on news recommendation do not incorporate visual informations, we could not locate previous research work that relies on the LXMERT model.

- **Research goal 2**:
  For RG2, we build a content-based fake news detection module, examine multimodal data, and provide relevant explanations utilizing explainable artificial intelligence. The experiments on two real-world datasets highlight the importance of learning the connection between two modalities. Hence, we have the following contributions:

(i) Elaborate a multimodal topic modeling analysis based on the Latent Dirichlet Allocation (LDA) topic model to measure the topic similarity between the text and the image within the online news content.
(ii) Analyze multimodal data within the news content to detect fake news based on image text alignment using Vision-and-Language BERT (VilBERT).
(iii) Generate appropriate multimodal explanations based on Local Interpretable Model-agnostic Explanations (LIME).
(iv) Implement and evaluate our system using two publicly available fake news datasets (i.e. Twitter and Weibo).

- **Research goal 3**:
  For RG3, we offer personalized awareness based on user profiles and especially centered around their reputation. Also, we provide users with explanations. Such clarifications are critical for reflecting the news' trustworthiness, raising user awareness, and ultimately influencing their behaviour in preserving both individuals' and society's security and privacy.

1.5. Thesis Structure

The rest of this thesis is organized as follows:

- Part I: General Context
  This part presents the general context of the thesis, and is composed of two chapters:
  - Chapter 1: Introduction presents an overlook of the context in which the project resides, its main motivations and objectives.
  - Chapter 2: Background and Related Work, discusses the recent research works regarding fake news detection methods and news recommendation systems.
- Part II: Methodology
  This section is dedicated to revealing the proposed system, and it is composed as follows:
  - Chapter 3: The Fake News Aware Recommender System, discusses the details of our proposed architecture. It highlights the main components of our proposed solution.
  - Chapter 4: Fake news Detection and Awareness, presents the two other components, EXMULF, fake news detection method and FNASY, the awareness component.
- Part III: Experiments
  This part details the experiments that were conducted:
  - Chapter 5: FNEWR: A New Dataset is presented.

  - Chapter 6: Experiments and Results, goes through the details of implementing the methods that were developed.
- Chapter 7: Conclusion, this final chapter concludes this thesis and discusses potential future research opportunities.

Chapter 2

Background and Related Work

Introduction

In this chapter, we present a summary of background and related work that is crucial to the understanding of our thesis. We examine previous efforts linked to each of the aforementioned research objectives, which may be divided into different areas, as illustrated in Figure 2.1.

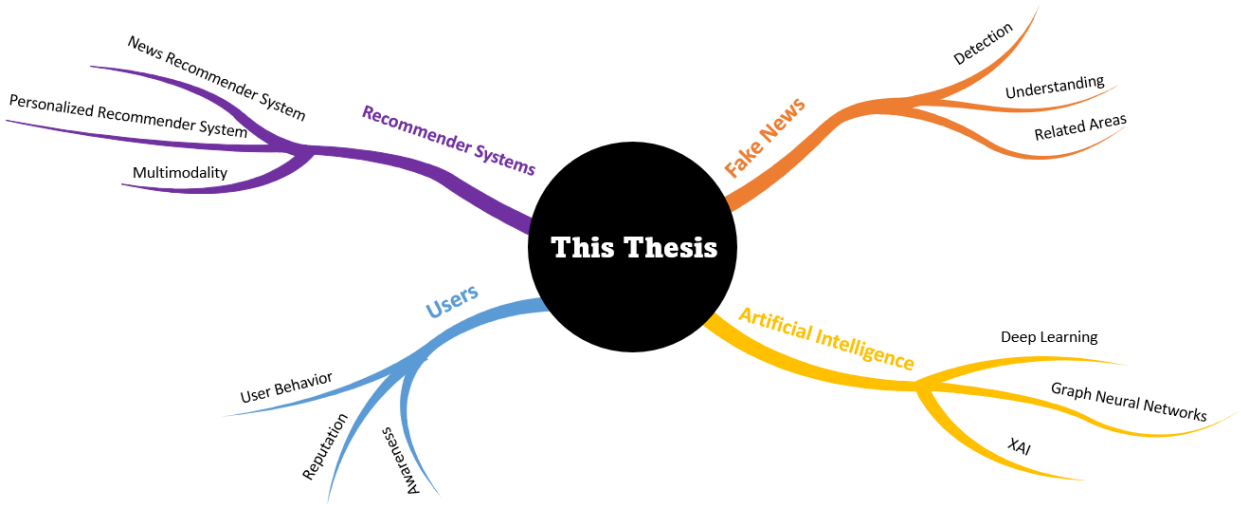


Fig. 2.1. Mind map of the themes related to this thesis.

2.1. Overview of Fake News

The expression "fake news" has grown immensely in popularity in recent years, crossing over from the sphere of social media users and into everyday language. The aim of this section is to explore the definition of fake news that is used in this thesis, as well as the terminology associated with it.

2.1.1. Fake News Definition

Fake news is still an issue without a clear or universally accepted definition. The term "fake news", according to the Collins English dictionary [**1**], is defined as "false, often sensational, information disseminated under the guise of news reporting". Many definitions were proposed in scientific literature. However, they vary based on the convergence and divergence of several related ideas from the provided definitions, such as satire, rumours, conspiracy theories, misinformation, and hoaxes.

The definition of this concept, as well as its interpretation, has increasingly been a source of debate [**18**]. As a result, it is critical to establish a baseline definition that will be used throughout this research. On that basis, we define "fake news" as follows:

Definition 2.1.1 (Fake News)**.** *A news article or message published and propagated through media, carrying false information regardless of the means and motives behind it.*

2.1.2. Related Areas

In this section, we will explore related areas of the fake news problem in order to illustrate some of the contrasts that arise.

Prior studies indicate that at least three types of fake news exist [**71**]. The first category is satire or parody, in which sites like the Onion[1] or Daily Mash[2] post fake news items in an attempt to criticize the media. The second category encompasses fake news that is somewhat real but utilized incorrectly, such as hoaxes, rumours, and misleading news that is not founded on facts but instead promotes an ongoing narrative. Finally, the third category includes news that was purposefully manufactured using misleading facts. Fake news is often created and spread on digital platforms with the intent of either making money via the number of clicks or causing confusion.

Hence, there is still no agreed-upon definition of the term "fake news." Furthermore, there are numerous terminologies and notions in literature that mention the concept of fake news. Table 2.1 defines several relevant keywords related to disinformation/misinformation/malinformation in order to provide a basic overview of the information disorder categories, including the main terms. Of course, the words in Table 2.1 are not all-inclusive.

[1] https://www.theonion.com/
[2] https://www.thedailymash.co.uk/

Table 2.1. A list of terms referring to different key aspects characterising information disorder [**72**].

Term	Definition
misinformation	unintentionally spread false information deceiving its recipients
malinformation	information that is partially or totally true, but spread with malicious intent
disinformation	intentionally spread and/or fabricated misinformation
fake news	disinformation in the format of news
hoax	disinformation that can have also humorous purposes
rumour	information that can be true and accurate, but still unverified; if it is falsified, it becomes misinformation
conspiracy theory	explanation of an event that assume a conspiracy by powerful group; a theory can make use of fake news, rumours as well as true information.
urban legends	kind of folklore made of rumours characterised by supernatural, horrifying or humorous elements
infodemic	mixture of misinformation and true information about the origins and alternative cures of a disease; expecially observed during COVID-19 pandemic
propaganda	malinformation that aims to influence an audience and a political agenda
click-bait	misinformation based strategy to deceive Web users and enticing them to follow a link
cherry-picking	malinformation practice that selects only the most beneficial information to the author's argument from what is available
hate speech	abusive malinformation that targets certain groups of people, expressing prejudice and threatening
cyberbullying	form of bullying that uses electronic communication, usually social media, that can contain misinformation, rumour and hate speech
troll	social media user that uses disinformation to increase the tension between different ideas
astroturf	disinformation practice of orchestrating a campaign masking its supporters and sponsors as grass-roots participants
crowdturf	crowdsourced astroturf
spam	unsolicited information that overloads its recipients
social bot	a social media user controlled by a software that mimic human behaviour; often used a tool for spamming, spreading misinformation, and astroturfing
satire	false information but intentionally harmless in the majority of cases, even if often it has strong political references and can be misused as a propaganda practice

Next our discovery of the fake news phenomena and its interconnected areas, we will investigate its origins in the following part.

2.1.3. From an age-old problem to a contemporary problem

Fake news are not a new phenomenon. According to Google Trends Analysis[3], the popularity of fake news peaked around the time of the 2016 US presidential election, however, the origins of fake news date back to before the printing press. Rumours and false stories have most likely existed as long as humans have lived in societies where power is shared. Until the advent of the printing press, information was mainly passed from person to person via word of mouth [**15**].

The history of fake news may be split into four divisions [**15**]. The earliest was the Pre–Printing Press Era, when news was written on materials such as stone and clay and was only available to the group's leaders. Some of the information glorified the leaders' majesty. As an example, we can cite the principal historian of Byzantium who used fake news to smear emperor Justinian. The second period is the Post–Printing Press Era. The invention of the printing press, together with the simultaneous growth of literacy, made it possible to disseminate information more broadly. For example, in his article "The Art of Political Lying" published in 1710, Jonathan Swift warns about political false news. Swift's comments on fake news in politics in 1710 are startlingly comparable to those of twenty-first-century authors. The third period is the Mass Media Era. Many reports about fake news have been published, the most notable was Orson Welles' broadcast The War of the Worlds in 1938. Finally, there is the Internet Era. The internet provides new ways for propagating fake news on a massive scale in the late twentieth century. In the early days of widespread internet use, some fake websites were launched. Some of these hoax websites were satirical, while others were designed to mislead or intentionally propagate biased or misleading news. As a result, the internet has transformed an age-old problem into a new menace[4].

2.2. Fake News Detection

Fake news detection is an ever-expanding research topic that is gaining a lot of attention since there are still a lot of challenges that need to be investigated. There are various research studies on fake news detection proposing different approaches. Relying on a survey [**111**], we sought to classify the different approaches into four categories, as shown in Figure 2.2.

The knowledge-based approach seeks to examine and/or detect fake news using a process known as fact-checking[5]. There are two kinds: manual fact-checking and automatic fact-checking. Investigating fake news from a style-based viewpoint emphasizes exploring the news content. The propagation-based perspective relies on facts related to the spread of fake

[3]https://trends.google.com/trends/?geo=CA

[4]https://theconversation.com/fake-news-the-internet-has-turned-an-age-old-problem-into-a-new-threat-72111

[5]Fact-checking aims to assess news authenticity by comparing the knowledge extracted from to-be-verified news content (e.g., its claims or statements) with known facts (i.e., true knowledge).

news, such as how it spreads and who spreads it. Finally, when researching fake news from a credibility standpoint, we examine news-related and social-related information. For example, a news piece published on an untrustworthy website(s) and sent by an untrustworthy user(s) is more likely to be fake news than news shared by authoritative and credible users.

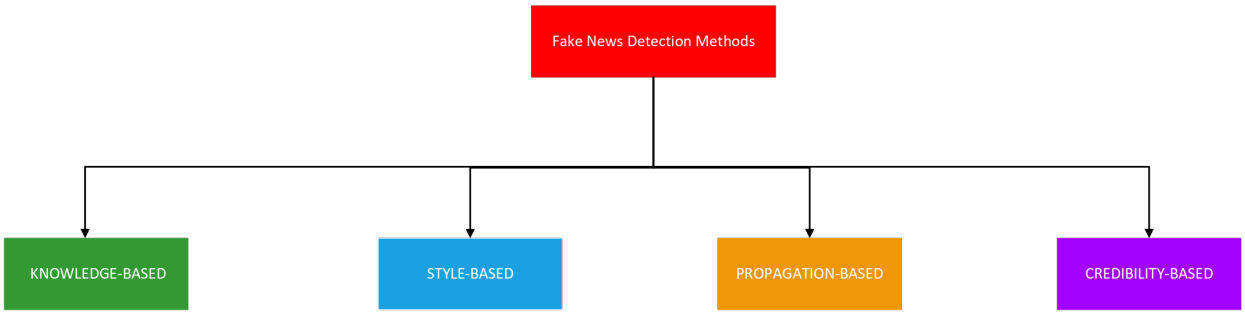


Fig. 2.2. Fake News Detection Methods.

This study includes a component that focuses on the detection of fake news based on content. As a result, in the following sections, we will cover related work on multimodal content-based fake news detection and explainable fake news detection.

2.2.1. Multimodal Content-based Fake News Detection

To date, various researchers in fake news detection have attempted to use visual information as auxiliary information in their detection algorithms to infer the veracity of online content. This is due to the fact that mixtures of these sources are commonly utilized to spread misinformation (e.g., a news title linked with an image from a different place, or from a different time)[**27**]. A comparison between these approaches with emphasis on the techniques and datasets used is provided in Table 2.2.

Some methods focus on the correlation between the attached images and the credibility of the news text [**98, 105, 108, 47, 57, 58, 32, 31, 79, 110, 64**], while others only use one or the other data type [**104, 84**].

Xue et al. [**98**] propose a neural network approach for fake news detection named MCNN (Multimodal Consistency Neural Network), using a similarity measurement module to measure the similarity of multimodal data (text and images).

Zeng et al. [**105**] define a fake news detection approach to comprehensively mine the semantic correlations between text-based content and the attached images.

Zhang et al. [**108**] propose an end-to-end model, named BERT-based domain adaptation neural network (BDANN) for multimodal fake news detection.

Kumari et al. [**47**] propose an attention-based multimodal fake news detection framework named AMFB, with a multimodal feature fusion, that leverages information from text and image and tries to maximize the correlation between them to get the most efficient multimodal shared representation.

Table 2.2. A comparison between the multimodal fake news detection approaches.

Reference	Techniques used	Datasets used
[**98**]	BERT [**22**], ResNet50 [**82**], cosine similarity.	MCG-FNeWS, PolitiFact, Twitter.
[**105**]	VGG model [**88**], multimodal variational autoencoder.	Twitter, Weibo.
[**108**]	BERT, VGG19.	Twitter, Weibo.
[**47**]	ABS-BiLSTM, ABM-CNN–RNN.	Twitter, Weibo.
[**57**]	VGG, Word2Vec, LSTM [**33**], cosine similarity.	Collected 1000 images from Google, Kaggle and onion for fake or real images with text.
[**58**]	Hierarchical Attention Network (HAN), Caption and Headline matching (CHM), Noise Variance Inconsistency (NVI), Error Level Analysis (ELA).	Fake News Detection by Jruvika, All Data, Fake News Sample by Guilherme Pontes.
[**32**]	BERT, VGG-16, cosine similarity.	FakeNewsNet.
[**31**]	Word2Vec, VGG19, LBP.	MediaEval, PolitiFact, GossipCop.
[**79**]	BERT,VGG19.	Twitter MediaEval, Weibo.
[**110**]	Text-CNN, Text-CNN, image2sentence, cosine similarity.	PolitiFact, GossipCop.
[**64**]	BERT, ResNet, attention mechanism.	Twitter, Weibo.
[**104**]	BERT, VGG19, Bi-LSTM, Graph-attention layer.	Twitter, Weibo.
[**84**]	Optical Character Recognition (OCR), Web scraping.	A dataset of thousands of images collected from Google Images, the Onion, and Kaggle.
[**75**]	Sentiment Analysis, Cultural Algorithms (CA).	Twitter, Weibo.

Mangal et al. [**57**] propose a fake news detection approach with the integration of embedded text cues and image features in which they extract text and objects available in the image and then check the similarity between them to find potential fraud in a piece of given information.

Meel et al.[**58**] propose a multimodal fake news detection framework that unitedly exploits hidden pattern extraction capabilities from text and visual image features.

Giachanou et al.[**32, 31**] propose multimodal, multi-image systems that combine information from different modalities (textual, visual and semantic information) in order to detect fake news posted online.

Singhal et al. [**79**] introduce a multimodal framework named SpotFake, which exploits both the textual and visual features of an article for fake news detection.

Zhou et al. [**110**] propose a similarity-aware fake news detection method named SAFE, which investigates multimodal (textual and visual) information to recognize the falsity of news articles based on their text, images, or their "mismatches".

Qian et al.[**64**] propose a hierarchical multimodal contextual attention network (HMCAN) for fake news detection by jointly modelling the multimodal context information (text and images) and the hierarchical semantics of text in a unified deep model.

Yuan et al. [**104**] propose an approach named DAGA-NN to improve fake news detection with a domain-adversarial and graph-attention neural network. However, their approach is based on a text forward environment with multiple events/domains instead of being based on multimodal data (text and image).

Vishwakarma et al. [**84**] propose an approach to detect the veracity of information on various social media platforms available in the form of images. The veracity of image text is validated by searching for it on web. Shah et al. [**75**] present a multimodal framework to detect fake news, without any further sub-task being considered, using a Cultural Algorithm with situational and normative knowledge.

2.2.2. Explainable Fake News Detection

Despite considerable advances in detecting fake news, minimal consideration has been devoted to explainability. Explanations focused on, machine learning have lately emerged as a promising path for achieving transparency in a variety of applications, including fake news detection.

Machine learning-based explanations aid in the clarification of a machine learning model's outcome. It provides explanations and illustrates the logic behind the resultant decisions and forecasting helps people understand how data is processed in order to come to a decision. For example, explainability in the detection of fake news entails describing why a certain piece of news was identified as fake news. This inhibits users from further spreading fraudulent material, thus limiting the detrimental impacts on both individual and societal security.

Multiple researchers [**76, 69, 99, 55, 63, 9, 21, 78**] are attempting to add predictability into their prediction models for fake news detection tasks. A comparison between these approaches with emphasis on the techniques and datasets used is provided in Table 2.3.

On the other hand, multiple studies on explainable machine learning are dedicated to investigating and evaluating existing fake news prediction models [**4, 48, 59**], including looking into which important features contribute to the models' prediction from the explainable machine learning perspective. For instance, Alharbi et al. [**4**] evaluate the trustworthiness of three existing fake news detection models (i.e. DEFEND, TCNNURG, and HSFD) through

Table 2.3. A comparison between the explainable fake news detection approaches.

Reference	Techniques used	Datasets used
[**76**]	Attention neural network.	PolitiFact, GossipCop.
[**69**]	SHAP.	BuzzFace.
[**99**]	MIMIC, ATTN, PERT.	An annotated benchmark dataset in the German language.
[**55**]	Co-Attention Network.	Twitter datasets: Twitter15, Twitter16.
[**63**]	Machine learning: linear method trained on stylometric features, a recurrent neural network method.	Fake News Corpus dataset.
[**9**]	Tsetlin Machine (TM).	PolitiFact, GossipCop.
[**21**]	NLP: semantic similarity and stance detection.	Clef18, FakeNewsNet, coinform250.
[**78**]	Network embedding learning.	PolitiFact, GossipCop.

the use of model-agnostic explainers (Captum, SHAP, and LIME,) in order to explain how the classification was made.

Kurasinski et al. [**48**] claim that there is a lack of research regarding the explainability of a machine learning-based fake news detection model, while efforts are mainly focused on its effectiveness. They investigate two classes of deep neural networks that are tasked with fake news detection (i.e. BiDir-LSTM-CNN, BERT), analyze them, and provide a deeper degree of explainability regarding the process. To explore how different types of explanations affect users in fake news detection, Mohseni et al. [**59**] designed four interpretable fake news detection algorithms (i.e. Bi-LSTM network, hierarchical attention network (HAN), Bi-LSTM teacher model with XG- Boost student model, BiLSTM network with Word2Vec word embedding). Their algorithms are dedicated to evaluating model explanations from multiple perspectives (i.e. user engagement, mental model, trust, and performance measurement). They report that adding explanations helped participants build appropriate mental models of the intelligent assistants in different conditions and adjust their levels of trust accordingly.

2.3. Personalized News Recommendation Systems

The general algorithms involved in recommender systems are categorized as being either collaborative filtering (CF), content-based filtering (CBF), or hybrid techniques. A CBF algorithm constructs a recommender by comparing the user profile and item profile based on the content of the shared attribute space. In contrast, the CF technique is content-free. CF takes advantage of user behaviour in terms of ratings, histories, and interactions with

objects. The hybrid filtering algorithms combine collaborative filtering with content-based filtering.

Deep learning-based solutions have emerged as a new branch of recommender systems in recent years. This is due to the fact that some deep neural recommender strengths, such as non-linear transformations, deep representations from input data, powerful modelling capability for sequential tasks, and improved capability of combining CF and CBF as hybrid models, can effectively address the limitations of conventional recommender systems [**66**].

A lot of research has been conducted by applying deep learning-based techniques for news recommendation systems (NRS). Convolutional neural networks (CNN) were employed in some research [**61, 106, 103**], knowledge-aware convolutional neural networks (KCNN) were used in others [**87**], and other scholars have investigated systems that incorporate different deep learning approaches. For example, in [**67**] the authors combined BERT and CNN, in [**112**], the suggested system associates LSTM and PCNN. Furthermore, several investigations have been done using the attention mechanism [**67, 112, 87, 106**].

The emphasis of this research is on a particular set of deep learning approaches, Graph Neural Networks (GNN), and multimodal recommendations. As a consequence, we will discuss related work on graph neural networks for NRS and multimodal NRS in the sections that follow.

2.3.1. Graph Neural Networks in News Recommender Systems

Among all deep learning algorithms, GNN is undoubtedly the most appealing approach due to its greater capacity to learn on graph-structured data, which is critical for recommender systems. For example, the interaction data in a recommendation application can be represented by a bipartite graph between user and item nodes, with observed interactions represented by links [**95**].

The Graph Enhanced Representation Learning (GERL) technique was proposed in [**30**]. The authors offer a news recommendation approach that can improve user and news representation learning by modelling their relatedness in a graph, in which users and news are both represented as nodes in a bipartite graph built from previous users click patterns. A transformer architecture is initially used to construct news semantic representations for news representations. Then, using a graph attention network, it is integrated with information from neighbor news in the graph. In terms of user representations, the researchers not only represent users from their previously clicked news, but they also combine the representations of their neighbor users into the graph with care. This method is comprised of a one-hop interaction learning module that represents the target user from previously clicked news and represents candidate news based on its textual content, and a two-hop graph learning module that uses a graph attention network to learn neighbour embeddings of news and users.

In [**38**], the authors introduce GNewsRec, a novel Graph Neural News Recommendation model that incorporates long-term and short-term user interest modelling. The approach fully exploits the high-order structural information between users and news items by first building a heterogeneous graph to describe the interactions and then using GNN to propagate the embeddings. GNewsRec is divided into 3 parts. The initial component of the process takes the news feature from the news headline and profile through CNN. The second component builds a heterogeneous user-news-topic graph from complete history of user clicks and uses GNN to encode high-order structure information for recommendation purposes. Learned user embeddings with complete previous user clicks are expected to convey reasonably consistent long-term user interests. In the third step, the researchers model the user's short-term interest with their recent reading history using an attention-based LSTM. Finally, for user representation, the user's long-term and short-term interests are combined.

In [**39**], the authors examine the high-order connectivity and latent preference factors underlying user-news interactions by suggesting a novel Graph Neural News Recommendation Model with Unsupervised Preference Disentanglement (GNUD). The model places user-news interactions on a bipartite graph and uses graph convolution to represent high-order relationships between users and news. Furthermore, a neighbourhood routing approach disentangles the learnt representations from various latent preference components, improving expressiveness and interpretability. A preference regularizer is also intended to drive each disentangled subspace to independently represent an isolated preference, thus boosting the quality of user and news embeddings.

A novel news recommendation mechanism is suggested in [**65**]. Interaction Graph Neural Network (IGNN) is a news recommendation model that incorporates a user-item interactions graph and a knowledge graph. Specifically, IGNN gets user and item representations using two graphs. One is the knowledge graph, while the other is the user-item interaction graph. It employs convolutional neural networks to learn content-based features at the knowledge and semantic levels, and a graph neural network to fuse high-order collaboration signals collected from the user-item interaction graph into the user and news representation learning process.

The article [**74**] offers a Multi-View Learning (MVL) framework for news recommendation that incorporates both the content and user-news interaction graph views. In the content view, the researchers employ a news encoder to learn news representations from various inputs such as titles, bodies, and categories. They obtain a representation of the user based on the candidate news article to be recommended from their browsed news. In the graph-view, the researchers suggest using a graph neural network to describe the interactions between various users and news to capture the user-news, user-user, and news-news relatedness in the user-news bipartite graphs. Furthermore, they propose incorporating an attention mechanism into the graph neural network in order to reflect the value of these interactions for more informative user and news representation learning.

The research [**43**] proposes the Temporal Sensitive Heterogeneous Graph Neural Network (TSHGNN), a time-sensitive heterogeneous graph neural network for news recommendation. The TSHGNN is composed of three major components: TCNN for news information extraction, Rein-LSTM for sequential feature extraction, and HANN for high-order structure information extraction. To extract news features from the news title, news entity, entity type, and active time, TCNN deploys a multi-channel convolutional neural network. The improved LSTM model is used by Rein-LSTM to extract the sequence aspects of the news that users click on. HANN builds a user-news-topic heterogeneous graph and uses a graph neural network to encode high-order structure information. Finally, the attention mechanism generates the corresponding user embeddings and candidate news embeddings. The similarity between the user feature representation and the candidate news indicates if the user clicks on the candidate news.

AGNN, an attention-based graph neural network news recommendation model, is proposed in [**42**]. Users, news, and topics are represented as three types of nodes in a heterogeneous network, with their interactions represented as edges. To produce corresponding news vectors, an attention based multi-channel CNN is employed. To extract the history of news clicks by users, an enhanced LSTM is deployed. These exploit the user-news topic heterogeneous graph to extract rich information. Meanwhile, they fuse the information of users' neighbors in the graph, i.e., news and topics, to ease the data sparsity problem. The attention mechanism then integrates this information with the user-clicked news.

A comparison between these approaches, based on news modeling, is provided in Table 2.4.

Table 2.4. Comparison of NRS models using GNN: News Modeling.

Reference	News Modeling	
	Information	Model
[**30**]	Title+Category+ User-News Graph	Transformer+GAT
[**38**]	Title+Entity+ Heterogeneous Graph	CNN+GNN
[**39**]	Title+Entity+ User-News Graph	CNN+Disentangled GCN
[**65**]	Title+Entity+ User-News Graph	KCNN+GNN
[**74**]	Title+Body+Category+ User-News Graph	CNN+Attention +GAT
[**43**]	Title+Entity+Entity type+Heterogeneous Graph	TCNN+GNN+Attention
[**42**]	Title+Entity+Entity type+Heterogeneous Graph	MCNN+AGNN

A comparison based on user modeling is provided in Table 2.5.

Table 2.5. Comparison of NRS models using GNN: User Modeling.

Reference	User Modeling	
	Information	Model
[**30**]	News Click+ User-News Graph	Self-Attention+GAT
[**38**]	News Click+ Heterogeneous Graph	LSTM+Attention+GNN
[**39**]	User-News Graph	Disentangled GCN
[**65**]	News Click+ User-News Graph	GNN
[**74**]	News Click+ User-News Graph	Self-Attention+GAT
[**43**]	News Click+ Heterogeneous Graph	HANN+Rein-LSTM +GNN+Attention
[**42**]	News Click+ Heterogeneous Graph	HANN+CLSTM +AGNN

2.3.2. Multi-modal News Recommender Systems

For news recommendations, accurate news representation is essential. The majority of current recommendation algorithms pay little attention to images in the news. In fact, images may be used to express news and forecast user behaviour. In fact, individuals may choose to read media stories not only because they are interested in the content of the news title, but also because they are attracted to the accompanying image [**8**].

The authors of [**93**] propose MM-Rec, a multimodal news recommendation system that incorporates both textual and visual news input to learn multimodal news representations. They begin by extracting regions-of-interest (ROIs) from news images using a pre-trained Mask R-CNN model for objective detection. Then, to acquire appropriate multimodal news representations, they employ a pre-trained visiolinguistic model, VilBERT, to encode both news texts and news image ROIs and model their inherent crossmodal relatedness using a co-attentional Transformer network. Furthermore, they offer a crossmodal candidate-aware attention network to pick relevant clicked news for user modelling by assessing the cross-modal relevance of candidate news and clicked news, which may assist in better modeling users' special interest in candidate news.

MRNT, a model that combines visuals and text in news, is suggested in [**100**]. To extract information from images, the Baidu Picture Recognizer is employed, while the Open Source Word Breaker is used to identify the language of news content. Following the completion of the vocabulary segmentation, news text tags are retrieved by deleting useless vocabulary such as stop words and function words. In this approach, new tags are derived from images and text in the news, and an adaptive tag (AT) algorithm is presented based on these new tags. Based on the user's feedback, the AT algorithm determines the tags which are of interest to the user.

Based on the user's preferences and multimodal content analysis, the authors suggest an implicit news recommender system, in [**23**]. The capacity to handle online press reports and

TV news feeds equally and concurrently is remarkable. To suggest relevant news articles to the user, they first model the user's Web behaviour by analysing the content of the RSS blogs to which the user has subscribed. This approach provides continually updated data in real time and allows for the tracking down of a specific user's personal profile by interpreting the user's blog post contents as a pretty reasonable estimate of their interests. On the basis of this profile, latent semantic analysis is utilized to find meaningful similarities and correlations between user-generated material and professional information items from online newspapers, press services, and television news providers. Finally, news items with strong connections to the user profile are recommended to the user. A comparison of different approaches is presented, with a focus on the methodology and datasets applied in Table 2.6.

Table 2.6. A comparison between Multimodal NRS approaches.

Reference	**Techniques used**	**Datasets used**
[**93**]	Mask R-CNN, VilBERT co-attentional Transformer.	Private dataset.
[**100**]	Correlation graph, Adaptive tag algorithm.	Private dataset.
[**23**]	Latent semantic analysis, NLP.	Private dataset.

2.4. Recommender Systems and Fake news

This section is devoted to works that mix the two domains of news recommendation algorithms and fake news. These efforts may be divided into two parts: recommender systems in favour of fake news and recommender systems that are anti-fake news.

2.4.1. RA: Culprit of misinformation spreading

Various factors promote the propagation of misinformation online, including how information is delivered, the digital platforms where it is distributed, and the algorithms that control suggestions of information within those platforms. While several studies have focused on the impact of various types of information, users, and digital platforms on the propagation of misinformation, there is a need to further investigate the influence of existing algorithms on the recommendation of inaccurate and misleading content since recommendation algorithms have been heavily criticized for being unintentional methods of amplification and diffusion of disinformation [**26**].

The authors in [**26**] investigated two well-known recommendation algorithms: collaborative filtering (CF) and content-based recommendation (CB). On the one hand, CF algorithms favor the repetition of the preference of the majority of users; these will most likely follow the trends (either spreading or avoiding misleading content) for issues where the majority of the community has already formed an opinion. CB algorithms, on the other hand, are

well-known for their portfolio effect and content overspecialization. As a result, it is expected that if a person consumes misleading material, these algorithms would reinforce this phenomenon.

In [**27**], the authors aimed to understand the impact of RAs on the formation of filter bubbles (where the only material users access is the type of content they like and that is generated by other individuals with similar opinions), as well as the influence of common popularity biases (i.e., the algorithm promotes information that is trending on the platform - e.g., getting more clicks) on the quality of items consumed by users. This study demonstrates how filter bubbles and popularity biases might make users more susceptible to misinformation by limiting the diversity and quality of information they are exposed to.

2.4.2. Recommender system to combat Fake news

Instead of being solely a part of the problem, recommendation algorithms might become a part of the solution. Many researchers have focused on how they may adjust RAs to avoid the spread of fake news. A comparison of the several techniques is given in Table 2.7.

Table 2.7. Comparison of RS methods to combat fake news.

Reference	RA	Avoid Fake news	Information used
[**62**]	CB	Calculate bias score .	URLs
[**56**]	Hybrid	Topic diversification	Title, image and content
[**102**]	CF	fact-checkers	URLs
[**85**]	CB	trustworthy guardians	URLs
[**52**]	CB	RL algorithm	headline and content
[**36**]	CB	TF-IDF vectorization.	URL and content.

To overcome this problem, the authors [**26**] suggested that users should be better profiled in order to capture their intentions and behaviours while distributing disinformation. They also offered recommendation systems that, rather than focusing on the concept of similarity between users and content, may be based on the concepts of similarity when it comes to users and dissimilarity when it comes to content.

Avoiding fake news has proven to be a significant challenge for social media platforms. Different online platforms are employing various tactics to minimize the spread of disinformation. Twitter, for example, started recommending popular tweets into the feeds of users who did not follow the accounts that posted them. This technique, of presenting popular contrasting viewpoints, was strongly criticized for promoting harsh political discourse and falsehoods [6].

[6]https://edition.cnn.com/2019/03/22/tech/twitter-algorithm-political-rhetoric/index.html

Other academics have focused on how recommendation systems may be used to prevent and reduce the spread of fake news. In [**62**], the authors suggest a technique for recommending relevant URLs from news sources with varying political biases on the same topic. They utilize a recent Pew research report to create a list of news sites that people of various political persuasions prefer to read. The writers then scrape news articles from various news sources on a variety of topics. Following that, they undertake a clustering approach to locate articles with related subjects, as well as to generate a bias score for each article. The article's bias score is calculated by adding up all of the biased sentences (those that include one or more terms from the bias lexicon) and dividing by the total number of sentences in the news article. In other words, they utilize the NPOV lexicon to determine whether sentences have more terms that are related to neutral words, and then use this information to compute the bias score. A Word2Vec model trained on the Wikipedia English whole article corpus is used to compute word similarity. Finally, for the current news article the user is reading, a bias score and additional articles on the same topic from previously gathered articles from other news sources are displayed.

In [**56**], the authors present a news recommender system prototype for gathering data from users, as well as an evaluation process for analyzing bubble development and false news interaction. To accomplish this, they used three collaborative-filtering recommendation algorithms: KNN user-based, KNN item-based, and matrix factorization with SVD, each of which was associated with one of the following post-filtering diversification strategies: Maximal Marginal Relevance (MMR) or Topic diversification (TD). The findings indicate that the diversification method can reduce the tendency to build bubbles and lower user involvement with fake news items.

In [**102**], the Attributed Multi-Relational Attention Network (AMRAN) was suggested. It is a deep-learning-based fact-checking URL personalized recommender system that relies on multi-relational context neighbors. The goal is to mitigate the negative impact of fake news on social media platforms such as Twitter and Facebook.

In [**85**], in order to further support credible information, the authors collect a large number of trustworthy guardians (those who correct disinformation and fake news in online discussions by linking to fact-checking URLs). In addition, they suggest a personalized recommender fact-checking URLs system to recommend similar URLs to guardians with similar interests. The goal is to urge the guardians to become more involved in fact-checking activities and to combat fake news. With that goal in mind, they present a novel URL recommendation model that takes advantage of fact-checking URL content, social network structure, and recent tweet content. GAU is a combination of the Guardian-Guardian SPPMI matrix, Auxiliary information (Modeling social structure, Modeling topical interests based on 200 recent tweets, and Modeling topical similarities of fact-checking pages), and the URL-URL SPPMI matrix.

In [**52**], the authors utilize a reinforcement learning (RL) model to learn how to implicitly redirect the user from fake news to real news. On top of a content-based recommender system, the RL algorithm learns a fake news intervention module. When a candidate fake news article is located, this intervention module is engaged to replace RS in recommending news and guiding the user to read the verified news.

In response to the URL or news article supplied by the users, the suggested technique in [**36**] employs a content-based recommendation system to offer verified news items to them. To select the most relevant terms that define the news storey, the scientists employed TF-IDF vectorization. After calculating the word importance, the calculation of relevance and similarity of one document with respect to other documents present in the dataset is calculated using the cosine similarity. The engine will produce the best suitable recommendation for the given input based on the similarity.

2.5. User Behavior and Fake news

Users are a critical component of the disinformation problem. As a result, several works have therefore studied the effect of different motivations and personalities and their effect on misinformation. These factors have been shown to impact users in the spread of disinformation, hence why it is critical to examine and record them as much as possible during the creation process of user-profiles for the purpose of recommendation [**26**].

Furthermore, we cannot rely solely on users ' good intentions; we must also implement an awareness system to educate users and minimize the spread of fake news. Thus, we focus on user awareness and user reputation in this section since we include these concepts in our suggested system. It is crucial to highlight that these areas haven't received much attention from researchers, and there hasn't been much study done on the subject.

2.5.1. User Reputation

The Crowd Signals methodology is one of the most often used methods for the automatic detection of Fake News. This method uses opinions (signals) expressed by a large number of users (crowd) to determine if a piece of news is fake or not. Although interesting, this strategy has a significant limitation: it is dependent on the user's stated opinion on the news, which is not always accessible. To solve this challenge, the paper [**28**] developed a Hybrid Crowd Signals (HCS), a technique based on crowd signals that incorporate implicit user views, deduced from the reputation (behaviour) of users towards the propagation of the examined news, rather than explicit ones, to detect Fake News.

The authors in [**73**] provided a study of the features (extracted from public data and metadata about users) in order to determine if and to what degree they are predictive of

social media user reliability. They also developed a deep learning-based architecture for predicting the class (reliable/unreliable) to be assigned to the user profile.

In [**101**], the researchers attempted to detect unsupervised fake news. They considered news as true and users' credibility as latent random variables, and exploited users' social media interactions to uncover their attitudes regarding the legitimacy of news. They employ a Bayesian network model to capture the conditional relationships between news facts, user opinions, and user trustworthiness.

2.5.2. User Awareness

On one hand, many academics have emphasized the need for fake news awareness in order to reduce the spread of fake news. For example, after researching the impact of false news on social media users in Nigeria, the authors, in [**6**], suggest that recommending fake news awareness methods is essential. They claim that fake news awareness encompasses consumers' understanding of the characteristics of fake news as well as their ability to recognize it.

The research, in [**35**], explores the theoretical foundations of trust-aware ranking in social recommenders. The authors believe that confirming the user-perceived quality of disruptive engagement is a critical frontier in developing such systems. They describe a trustworthy recommender as a recommender that enables transparent and interpretable interaction against the rising current of dogmatization and partisan antipathy. The authors discovered that the constituent characteristics of trustworthiness (diversity, transparency, explainability, and disruption) open up new avenues for preventing dogmatism and developing decision-aware, transparent news recommender systems.

Other studies, on the other hand, investigate the level of consciousness. For example, in [**29**] the researchers suggest an Awareness Index to compute the knowledge degree and the truthfulness of the news of the 293 volunteers in order to examine people's awareness. The purpose of the Awareness Index is to weigh every single viewpoint and offer a level of knowledge that participants have on each particular piece of news. They created a 7-point psychometric Likert scale (Strongly Agree, Agree, Somewhat Agree, Neither Agree nor Disagree, Somewhat Disagree, Disagree, Strongly Disagree) and invited volunteers to express their opinions in order to better understand how people interpret every item of news.

Other researchers incorporate the awareness component into their suggested system such as in [**62**], where the authors inform the reader how biased the news article is by calculating and displaying a bias score to benchmark the viewed article and suggest other articles on the same topic from different news sources, allowing the user to make a decision on what they want to read.

2.6. Research Gaps

Despite various efforts to develop solutions to fake news detection as well as news recommendation systems, there are still opportunities for advancements. Based on that, we highlight the research gaps that inspired our work:

- The majority of present initiatives are either focused on fake news detection [**105, 108, 47**] or on news recommendation systems [**30, 38, 39**], but neither can truly tackle the disinformation problem. On the one hand, spotting fake news without acting on it cannot generate any advancement in terms of user behaviour. On the other hand, enhancements in recommendation algorithms that provide strong results in terms of personalization and accuracy do not always imply that the system will supply users with the right/real content. Table 2.8 illustrates a comparison of our system to the most advanced ones, with a focus on the news recommendation algorithm, fake news detection and information utilized for that, and user awareness factors. In this study, we propose to adapt recommendation algorithms in a way that recommendations are solely based on trustworthy users.

Table 2.8. Overview of the state-of-the-art methods Vs our approach.

Reference	NRS	Fake news		User awareness
		Detection methods	Information used	
[**62**]	CB	Calculate bias score .	URLs.	
[**56**]	Hybrid	Topic diversification.	**Multimodal**.	
[**102**]	CF	Fact-checkers.	URLs.	
[**85**]	CB	Trustworthy guardians.	URLs.	
[**52**]	CB	RL algorithm.	Content.	
[**36**]	CB	TF-IDF vectorization.	URL.	
Our approach	**New algorithm**	**EXMULF**	**Multimodal**	✓

In the upcoming chapters, we will go through the news algorithm and EXMULF component.

Our effort integrates the two domains, respectively, fake news detection and news recommendation, in order to limit the spread of misinformation and make users more aware of it. As a result, this initiative covers not just fake news detection or NRS, but also user awareness and behavioural changes.

- We couldn't come across any studies that look into recommendation algorithms. Previous research employed existing algorithms, as seen in the table 2.2, and simply tried to propose an alternative if something was deemed to be fake news. There has been no inquiry into how we may develop a new algorithm to aid in countering the

spread of disinformation. In our approach, we modified a well-known algorithm in order to achieve our goals.

- Previous attempts in the area of fake news proposed multimodal content detection as a viable solution to handle the fake news problem on digital platforms. However, explainability can provide users with further clarifications. Table 2.9 below offers a comparison between our system and state-of-the-art ones in fake news detection, with an emphasis on multimodality, explainability, and news content aspects (i.e. whether the detection is based only on the news content).

Table 2.9. Overview of the state-of-the-art methods for fake news detection.

Approach	**Multimodal**	**Explainable**	**News content**
Shu et al. [**76**]		✓	
Reis et al. [**69**]		✓	
Yang et al. [**99**]		✓	
Lu et al. [**55**]		✓	
Przybyła et al. [**63**]		✓	✓
Bhattarai et al. [**9**]		✓	✓
Denaux et al. [**21**]		✓	✓
Silva et al. [**78**]		✓	
Xue et al. [**98**]	✓		✓
Zeng et al. [**105**]	✓		✓
Zhang et al. [**108**]	✓		✓
Kumari et al. [**47**]	✓		✓
Mangal et al. [**57**]	✓		✓
Meel et al. [**58**]	✓		✓
Giachanou et al. [**32**]	✓		✓
Giachanou et al. [**31**]	✓		✓
Singhal et al. [**79**]	✓		✓
Zhou et al. [**110**]	✓		✓
Qian et al. [**64**]	✓		
Yuan et al. [**104**]	✓		
Vishwakarma et al. [**84**]	✓		✓
Shah et al. [**75**]	✓		✓
EXMULF	✓	✓	✓

In the upcoming chapters, we will go through the EXMULF component.

In our approach, we take into account all factors. We adopt VilBERT (Vision-and-Language BERT) multimodal alignment for a multimodal content-based fake news detection in which we predict if the related picture is aligned with the text of the news body. In fact, VilBERT is pretrained on the conceptual captions dataset using two training objectives, masked multimodal learning and image text alignment prediction.

The latter is what motivates us the most to employ VilBERT in our multimodal detector component. We choose to use ViLBERT because of its high performance on a variety of visiolinguistic tasks, including visual question answering and image retrieval. Explainable AI (XAI) for multimodal explainable content-based fake news detection employing LIME (Local Interpretable Model-agnostic Explanations).

- Last, the vast majority of awareness studies emphasize the need of being aware of fake news rather than offering ways to enhance individual awareness. However, with our method, we provide a personalized awareness component to guide the user to the actual news. This component will notify the user whether the news is phoney or true and will provide explanations based on prior news that he has clicked on.

Conclusion

In this chapter, we reviewed related works and the background knowledge required to understand this thesis, highlighting the fake news ecosystem. In addition to defining fake news, we presented current approaches proposed to tackle this problem. We also highlighted some characteristics of recommender systems that drive the propagation of fake news. Moreover, we described user reputation and user awareness. Last, we presented the research gaps addressed in this thesis.

The next part will offer a detailed description of the proposed method. Additionally, we present the different steps taken towards achieving the goals of our project.

Part II

Methodology

This part is dedicated to the proposed methodology. It begins by providing a general overview and outlining the overall workflow chosen. Next, in the upcoming chapters, the details of the various components are explored in-depth in order to demonstrate their functionalities.

The primary goal of this work is not only to detect fake news but also to reduce their propagation. To achieve this purpose, we design and implement a modularized system capable of recommending news to the user while detecting fake news and helping users to be more aware of this issue.

The system is primarily comprised of three modules:

- **Fake News Aware Recommender system (FANAR)**: provides personalized news recommendations and aids in the reduction of fake news spread.
- **Explainable Multimodal Content-based Fake News Detection System (EXMULF)**: classifies the recommended news and generates explanations.
- **Fake News Awareness System (FNASY)**: tries to educate consumers about fake news.

In order to better describe the proposed architecture, we provide the following overview of the entire system, as shown by illustration 2.3 below.

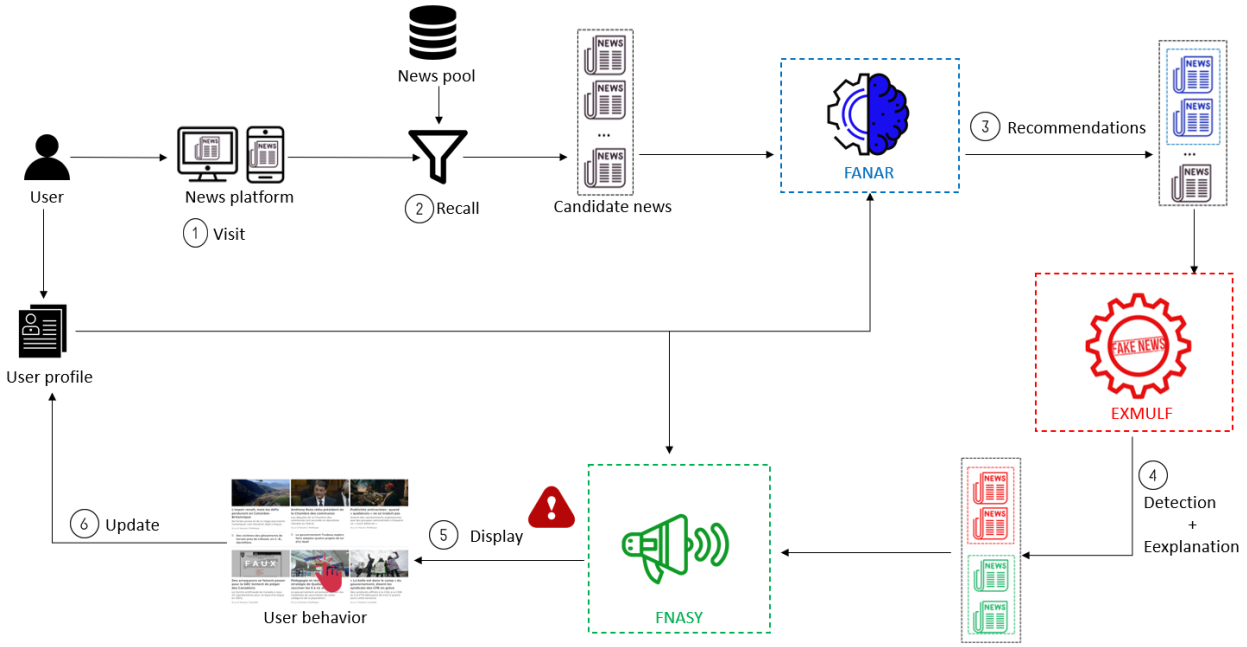


Fig. 2.3. The system workflow.

When users enter a news platform, the platform will retrieve a small set of candidate news from a large-scale news pool, and **FANAR** (FAke News Aware Recommender system), the personalized news recommender, will provide some recommendations to individual users based on their preferences as determined by their profiles. The recommended news will then be fed into **EXMULF** (Explainable Multimodal Content-based Fake News Detection System), which will provide the predicted class (fake/real) of the item as well as pertinent explanations. Following that, the **FNASY** (Fake News Awareness System) will offer customized nudges with the recommended news depending on the user profile, specifically user reliability. Finally, the platform will collect the user's actions towards these news so as to update the kept user profile for future services.

We propose a series of adaptations in the system in order to tackle the fake news problem. these adaptations are in two different steps in the workflow. First, we have an adaptation that takes place within the recommendation. FANAR is mainly a new recommendation algorithm that aids in the reduction of fake news spread by avoiding unreliable neighbors. Second, we have a post recommendation adaptation. In fact, the general system doesn't just detect fake news using EXMULF, but also tries to educate consumers about them.

It is essential to note that an obvious solution would be to simply exclude fake news from the recommendation set. However, under the current system, the consumer has complete control over whether or not he consumes fake news. Our solution does not remove fake news from the system, instead, we allow the user to choose whether or not to consume it, because limiting user choice is equivalent to censorship. To combat misinformation, we rely on the system's awareness part, which will notify and explain to the end-user that a given news is indeed fake and should be avoided.

To summarize, the proposed system consists mainly of three components: FANAR, EXMULF, and the awareness component. Following that, we will go through each component in depth.

Chapter 3

The Fake News Aware Recommender System

Introduction

This chapter introduces FANAR, **FA**ke **N**ews **A**ware **R**ecommender system. The system presents a novel technique to efficiently and effectively recommend news to users while avoiding the problem of fake news. We also look at the user reputation issue in order to calculate user reliability.

The chapter is organized as follows: First, we provide an overview of the system, emphasizing the reasoning behind our various decisions and the problem formulation. The model is then shown with details of each of its component. Finally, we define user reputation and offer a probabilistic technique for estimating it.

3.1. General Overview

Although there are significant advantages of news recommendation systems and fake news detection, designing an effective system capable of detecting, recommending, and assisting users still necessitates considerable work. Thus, we attempted to include all of these aspects into this work. This chapter focuses on the news recommendation component. We attempt to create a news recommendation algorithm that is not only capable of providing personalized recommendations but is also aware of fake news.

In the recommendation process, we chose to employ Graph Neural Networks. Their core concept is to use neural networks to iteratively gather feature information from local graph neighbors. Meanwhile, following transformation and aggregation, node information may be propagated across a graph. As a consequence, GNNs naturally incorporate node information as well as a topological structure and have been shown to be effective in learning from graph formation and topological structure data. Because data in social recommender systems may be represented as a user-user social graph and a user-item graph, and learning latent variables of users and things is the key, GNNs have a huge potential to advance social recommendation.

Most existing news representation algorithms simply learn news representations from news texts, neglecting visual information in news such as images. In reality, users may choose to read news stories not just because they are interested in the news's content, but also because they are captivated by the associated image. As a result, the visual information in news images can give valuable additional information for news comprehension and user action prediction. We investigate ways to utilize visual news information to improve news recommendations in this work. To learn multimodal news representations, we offer a multimodal news recommendation approach that incorporates both textual and visual news information.

This module, as shown in figure 3.1 below, takes in a news candidate and a user profile and delivers personalised recommendations.

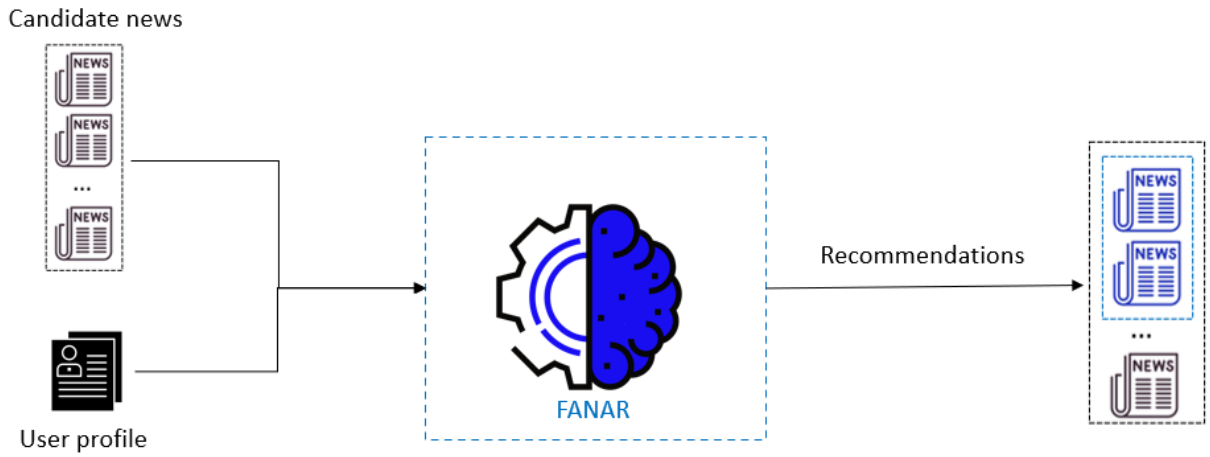


Fig. 3.1. FANAR Overview.

Our main contributions regarding the recommendation part of our system are listed as follows:

- To the best of our knowledge, our work is the first to include the concept of fake news into a recommendation engine.
- We provide a novel collaborative filtering strategy for the news recommendation assignment that may considerably prevent the propagation of fake news by avoiding untrustworthy neighbors.
- In this part, we provide a new probabilistic model for user reputation in the context of fake news.
- We propose a multimodal way of representing news. Although the majority of the studies on news recommendation do not incorporate visual pieces of information, we couldn't locate previous works that use the LXMERT model.

3.2. Preliminary and Problem Formulation

This section will first provide some introductory terminologies as utilized in this project, followed by a description of the problem formulation.

3.2.1. Preliminary

- **Sequential Recommendation**[**107**]:
  In the setting of sequential recommendation, let $\mathbb{U}$ and $\mathbb{I}$ represent the set of users and items, respectively. For each user $u \in \mathbb{U}$, its action sequence is denoted as $\mathbb{S}^u = (i_1, i_2, \ldots, i_k)$, where $i \in \mathbb{I}$, $\mathbb{T}^u = (t_1, t_2, \ldots, t_k)$ is the corresponding timestamp sequence of $\mathbb{S}^u$. The set of all $\mathbb{S}^u$ is denoted as $\mathbb{S}$. The object of sequential recommendation is to predict the next item of $\mathbb{S}^u$ employing sequence information before time t_k.
- **Dynamic Graph**[**107**]:
  A dynamic network can be defined as $\mathbb{G} = (\mathbb{V}, \mathbb{E}, \mathbb{T})$, where $\mathbb{V} = (v_1, v_2, \ldots, v_n)$ is the node set and $e \in \mathbb{E}$ represents the interaction between v_i and v_j at time $t \in \mathbb{T}$, so edge e_{ij} between v_i and v_j is generally represented by triplet (v_i, v_j, t). In some cases, t can also indicate the order of interactions between two nodes. By recording the time or order of each edge, a dynamic graph can capture the evolution of the relationship between nodes.
- **Dynamic Recommendation**[**49**]:
  Let $\mathbb{U}$, $\mathbb{V}$ represent the user and item sets, respectively. In a dynamic recommendation scenario, the i-th user-item interaction is represented in a tuple $\mathbb{S}_i = (u_i, v_i, t_i, f_i)$, where $i \in \{1,2, \ldots, \mathbb{I}\}$, and $\mathbb{I}$ is the total number of interactions. $u_i \in \mathbb{U}$, $v_i \in \mathbb{V}$ are the user and item in the interaction and t_i is the time stamp. f_i denotes features of the interaction, and it includes user features f_u and item features f_v. The target of dynamic recommendation is to learn the representations of the user and item from current interaction and historical records, and then predict the most possible item that the user will interact with in the future

3.2.2. Problem Formulation

The concept of news recommendation can be articulated as follows. We want to anticipate if a user u_i will click a candidate news $n_{candidate}$ that it hasn't seen previously, based on the user-news history interactions (U, N)) and neighbors. Table 3.1 summarizes all of the notations.

Let $U \in \{u_1, u_2, \ldots, u_n,\}$ and $N \in \{n_1, n_2, \ldots, n_m,\}$ be the sets of users and news respectively, where n is the number of users, and m is the number of news.

Table 3.1. Notations.

Symbols	Definitions and Descriptions
q_j	The embedding of news n_j.
p_i	The embedding of user u_i.
d	The length of embedding vector.
$Q(i)$	The set of news which user u_i interacted with.
$P(i)$	The set of neighbors of the user u_i.
$K(i)$	The set of users who have interacted the news n_j.
R	The user-news rating matrix (user-news graph).
W, b	The weight and bias in neural network.
h_i	user u_i's latent factor.
h_i^N	The news-space user latent factor.
h_i^S	The neighbor-space user latent factor.
$Z(i)$	$\subseteq Q(i)$ The set of the real news that user u_i has interacted with.
$T(i)$	User u_i's reliable neighbours.
$Z(i)$	$Z(i) \subseteq Q(i)$ is the set of **the real news** that user u_i has interacted with.
x_{iz}	A representation vector to denote interaction between u_i and an item n_z.
Agg_{news}	The aggregation function.
σ	Non-linear activation function (i.e., a rectified linear unit).
$Agg_{reliable-neighbors}$	The aggregation function on user's reliable neighbors.
l	The index of a hidden layer.

We assume that $M \in R^{n \times m}$ is the user-news interaction matrix, which is also called the user-news graph. If read n_j, $m_{ij} = 1$, otherwise $m_{ij} = 0$. Let $P(i)$ be the set of neighbors (i.e. users) of the user u_i , $Q(i)$ be the set of news which u_i have interacted with. We denote the embedding vectors $p_i, q_j \in \mathbb{R}^d$ to represent, respectively, user u_i and news n_j, where d is the length of the embedding vector.

3.3. FANAR Model Architecture

Figure 3.2 illustrates the envisaged model's architecture. The model is divided into three parts: user modelling, news modelling, and click prediction.

The first component is user modelling, which is used to learn users' latent variables. We have a great way to explain user representations from several angles as we introduce two different graphs, namely a neighbors graph and a user-news graph. As a result, two types of aggregations are established to process these two separate graphs. One is news aggregation, which may be used to understand users through interactions between users and news in the user-news graph (or news-space). The other is neighbors aggregation, which is the relationship between users in the neighbors network (or neighbor-space). Then, it is natural to get user latent factors by combining information from both the item space and the social space. The second component is news modelling, which is used to represent the

news multimodal content. In this component, we employ a visiolinguistic model, named LXMERT, to illustrate the cross-modality between the text and the image of the news. The third component involves learning model parameters through prediction by combining user and news modelling components.

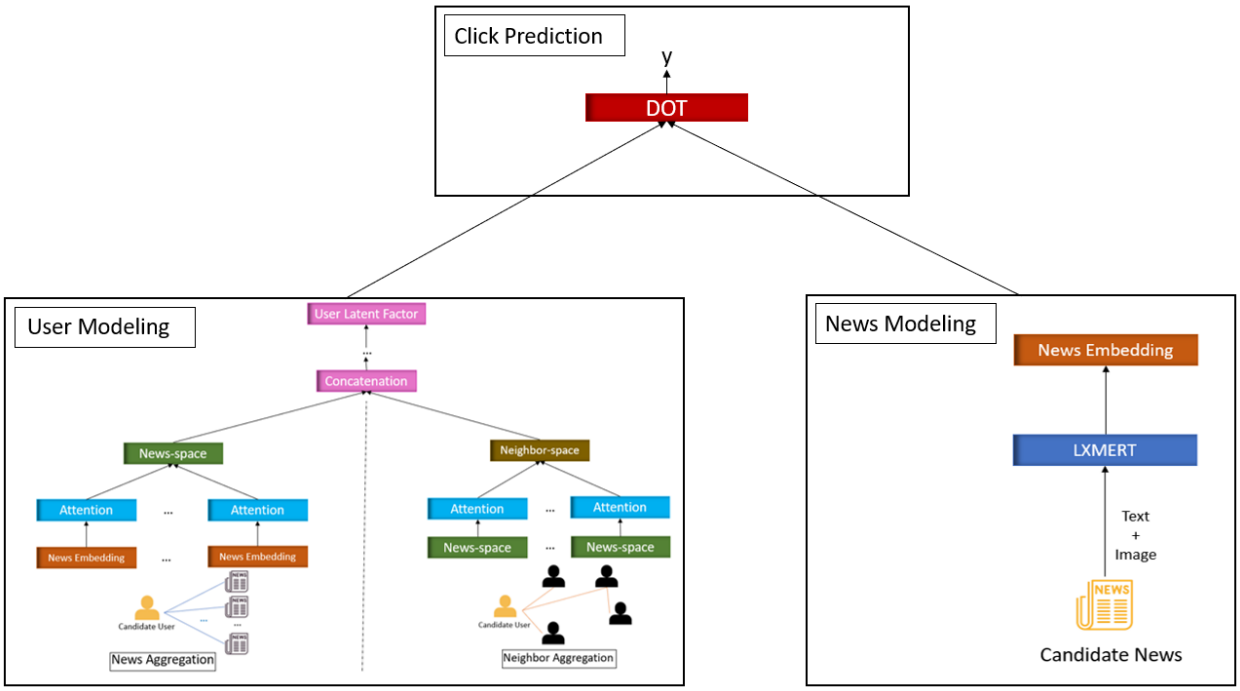


Fig. 3.2. FANAR Model.

Following that, we will go through each model component in further detail.

3.3.1. News Modeling

Users may choose news not only because of their interest in the text but also because of the attractiveness of the news images. Thus, representing the visual content of news, such as images, is critical for learning news representations. The right portion of Figure 3.2 illustrates the architecture of the multimodal news modelling. It requires an image and the text of a news article as input. We denote news representation by z_j.

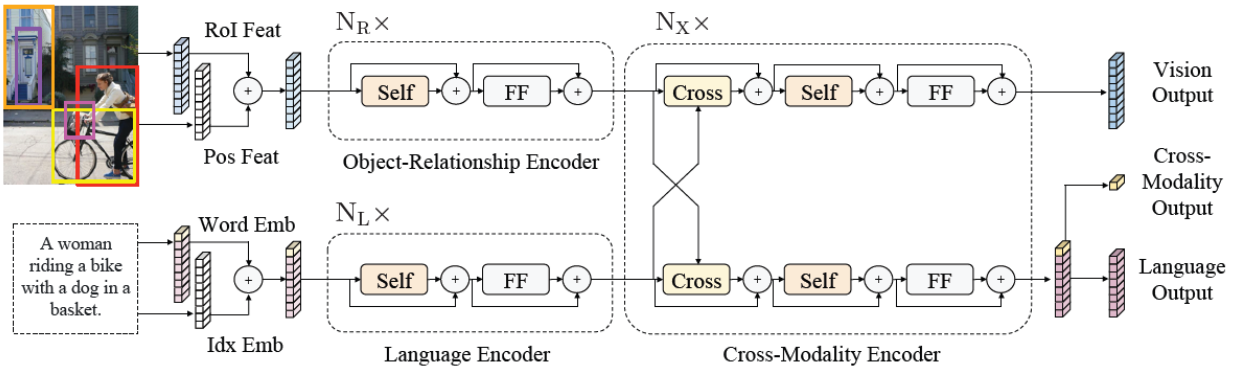


Fig. 3.3. The LXMERT model for learning vision-and-language cross-modality representations [**81**].

Modelling news texts and images separately is an intuitive method. However, the text and image of the same news item are frequently related in some way. Capturing the relationship between news text and images might help us better comprehend their content and predict user preferences. Visiolinguistic models are useful for simulating the crossmodal relationships between texts and visuals. As a result, we suggest using a pre-trained visiolinguistic model called **LXMERT** [**81**] to capture the cross-modality between news text and image while learning representations of both. The name stands for **L**earning **Cross-M**odality **E**ncoder **R**epresentations from **T**ransformers.

The model, as illustrated in Figure 3.3, receives two inputs: an image and its associated text. Each image is represented by a series of objects, and each sentence by a set of words. The model can create language representations, image representations, and cross-modality representations from the inputs due to the combination of self-attention and cross-attention layers. According to the authors, their model consists of three encoders: an object relationship encoder, a language encoder, and a cross-modality encoder.

The authors extract areas from an image for image preprocessing. They accomplish this by using a Faster R-CNN model with a ResNet-101. Each region is represented by its position and features of interest to that region. Feedforward layers are used to learn the representation of each object. To start, the term **language encoder** refers to the N_L transformer blocks. The authors do not employ a pre-trained BERT model for it, and they even show that using a pre-trained BERT model will significantly damage the outcomes. Thus, the layers are randomly initialized, with the hidden state size set to 768 and N_L set to 9. Second, N_R transformer blocks are referred to as **object relationship encoder**. The layers are also initialised at random, with the hidden state size set to 768 and N_R set to 5. Finally, the **cross-modality encoder** incorporates transformer layers preceded by co-attentional transformer layers in which the language stream's keys and values are the visual

stream's query, and vice versa. The hidden state size is at 768 and N_X at 5 and again randomly assigned to the layers.

To summarize, we suggest using a visiolinguistic model to encode both news texts and images while also capturing their essential crossmodal relatedness. While most existing news recommendation systems disregard image-related information, our method combines both textual and visual news information into news representation learning while modelling their intrinsic relatedness for enhanced news content comprehension.

3.3.2. User Modeling

The aim of user modelling is to identify a user's latent factors, denoted by h_i. The difficulty is figuring out how to effectively merge the user news graph with the neighbour graph. To overcome this obstacle, we first employ two methods of aggregation to learn factors from two graphs, as illustrated in Figure 3.2 on the left side. The first type of aggregation, news aggregation, is used to learn the news-space user latent, h_i^N, from the user-news graph. The second type of aggregation is neighbour aggregation, which uses the neighbour graph to learn the neighbor-space user latent factor, h_i^S. These two components are then merged to generate the final user latent factors.

3.3.2.1. News Aggregation.

We present a strategy meant to capture interactions in the user-news graph for learning news-space user latent factor h_i^N, which is utilized to model user latent factors via interactions in the user-news graph. The goal of news aggregation is to understand news-space user latent factors h_i^N, by taking into account news that a user has engaged with.

We use the following function to illustrate this aggregation mathematically:

$$h_i^N = \sigma(W.Agg_{news}(\{x_{iq}, \forall z \in Z(i)\}) + b)$$

where:

- $Z(i) \subseteq Q(i)$ is the set of **the real news** that user u_i has interacted with,
- x_{iz} is a representation vector to denote interaction between u_i and an item n_z,
- Agg_{news} is the aggregation function,
- σ denotes non-linear activation function (i.e., a rectified linear unit),
- and W and b are the weight and bias of a neural network.

One popular aggregation function for Agg_{news} is the mean operator where we take the element-wise mean of the vectors in $\{x_{iz}, \forall z \in Z(i)\}$.

$$h_i^N = \sigma(W.(\sum_{z \in Z(i)} \alpha_i x_{iz}) + b)$$

where α_i is fixed to $\frac{1}{|Z(i)|}$. However, this may not be optimal, due to the fact that the influence of interactions on users may vary dramatically. As a result, we should allow interactions to contribute deferentially to a user's latent factor by attributing a specific weight to each interaction, inspired by [**16**], an effective path is to modify α_i such that it is aware of the target user u_i, i.e., providing an individualized weight to each (n_z, u_i) pair :

$$h_i^N = \sigma(W.(\sum_{z \in Z(i)} \alpha_{iz} x_{iz}) + b)$$

Following the work presented in [**16**],

$$\alpha_{iz} = \frac{exp(\alpha_{iz}^*)}{\sum_{z \in Z(i)} \alpha_{iz}^*}$$

where the attention network formally defined as [**16**]

$$\alpha_{iz}^* = w_2^T \sigma(W_1[x_{iz} \oplus p_i] + b_1) + b_2$$

α_{iz} denotes the attention weight of the interaction with n_q in contributing to user u_i's news-space latent factor when characterizing user u_i's preference from the interaction history $Z(i)$.

3.3.2.2. Neighbour Aggregation.

Only the **most reliable neighbors** are taken into account in this case $T(i) = u \in P(i)$ where u is reliable. Actually, the neighbour graph is a **dynamic graph** since the user reliability is updated as the user consumes additional news. As a result, a trustworthy user at time t may become untrustworthy at time $t+n$ if he attempts to consume more fake news.

Before we proceed any further, it is important to note that in this study we computer user reliability/reputation. In the second section of this chapter, we cover user reputation in-depth. Furthermore, we observe that the term neighbourhood is employed in the same way in classical recommendation systems to represent users chosen based on similarity, and it has nothing to do with the structure of the social network.

Instead, the neighbourhood is defined as the group of K reliable users who have more clicked news in common with the candidate user. To compute the similarity, we utilize the **Jaccard similarity coefficient** between user u_i and user u_j as follows:

$$\mathbb{J}_{ij} = \frac{|Q(i) \cap Q(j)|}{|Q(i) \cup Q(j)|}$$

Where $Q(i)$ and $Q(j)$ are the sets of clicked news by user u_i and user u_j respectively.

In order to represent user latent factors from this social perspective, we propose neighbor-space user latent factors, which is to aggregate the news-space user latent factors of reliable

neighboring users from the neighbors graph. Specifically, the neighbor-space user latent factor of u_i, h_i^S, is to aggregate the news-space user latent factors of users in u_i's neighbors $T(i)$, as the follows:

$$h_i^S = \sigma(W.Agg_{reliable-neighbors}(\{h_r^N, \forall r \in T(i)\}) + b)$$

where $Agg_{reliable-neighbors}$ denotes the aggregation function on user's reliable neighbors.

Then we proceed in the same way as we did with news aggregation:

$$h_i^S = \sigma(W.(\sum_{r \in T(i)} \beta_{ir}.h_r^N) + b)$$

Since the neighbour graph and the user-news graph give information about users from various viewpoints. We suggest combining these two latent components into the final user latent factor using a typical MLP in which the news-space user latent factor and the neighbor-space user latent factor are concatenated before input into the MLP.

$$g_1 = (h_i^N \oplus h_i^S)$$
$$g_2 = \sigma(W_2.g_1 + b_2)$$
$$\dots$$
$$h_i = \sigma(W_l.g_{l-1} + b_l)$$

where l is the index of a hidden layer.

3.3.3. Recommendation and Model Training

For a user u_i along with his explored news and a candidate news item n_j to be recommended, our goal is to predict the probability of the user browsing that candidate news. The final representations of user h_i and news z_j. The click probability score y is calculated by the inner product of the representation vectors of the user and the candidate news as $y = h_i.z_j$.

Inspired by [**93**], we employ negative sampling strategies, for model training. We randomly choose K news articles that are not clicked by this user as negative samples for each news article explored by a user that is regarded as a positive sample. Then, we jointly predict the click probability scores of the positive news y^+ and the K negative news $[y_1^-, y_2^-, \dots, y_K^-]$. In this way, we formulate the news click prediction problem as a pseudo $K + 1way$ classification task. We normalize these click probability scores using softmax to compute the posterior click probability of a positive sample as follows:

$$p_i = \frac{exp(y_i^+)}{exp(y_i^+) + \sum_{j=1}^{k} exp(y_{i,j}^-)}$$

where y_i^+ is the click probability score of the i^{th} positive news, and $y_{i,j}^-$ is the click probability score of the j^{th} negative news in the same session with the i^{th} positive news.

The loss function for model training is the negative log-likelihood of all positive samples, which can be formulated as:

$$L = -\sum_{i=1}^{S} log(p_i)$$

where S is the size of the set of the positive training samples.

3.4. User Reputation

In this section, we will explore the user reputation. To start, we define user reliability. Then we proceed to the calculation module.

3.4.1. Overview

The reputation and influence of a person on social media is a new topic of study that is gaining attention. Knowing the reputation and influence of users, as well as being able to predict both, may be beneficial in many sectors of business, including viral marketing, information broadcasting, recommendation systems, searching, and social customer relationship management, to name a few [**3**]. In our circumstance, we need to know about user reputation in order to avoid untrustworthy neighbors throughout the recommendation process.

It is essential to note that reputation is a very complicated subject that is influenced not just by users' behaviour but also by a variety of other factors such as their intentions. As a result, we will first attempt to define user reputation. Then, we talk about how we can compute it.

3.4.2. User Reputation: Definition

The reputation of users in terms of their ability to spot Fake News is typically based on explicit user opinion, which is not always available. As a result, we may assess the implicit opinions deduced from user behaviour on the dissemination of the investigated news. This deduction is founded on the notion that when a person disseminates news via digital media, they want to indicate, whether maliciously or not, that they believe the news to be genuine. The fact that user u posts the news n is an implied indication that n is not fake in u's viewpoint. As a result, when a person decides to distribute n in digital media, whether maliciously or not, they want to demonstrate to other users in digital media that they believe n to be real [**28**].

This concept is captured by a quotation from philosopher **Habermas** [**28**], who states that every communicative action has an inevitable claim to truth. As a result, u's disclosure

of n is a signal (i.e., implicit opinion) that u wishes to externalize the fact that n is not a fake. The reputation of these users is determined based on their hits or misses in providing implicit judgments about news that have already been reviewed and whose labels are known [**28**]. Hence, unlike the explicit signals-based method for calculating user reliability, we do not require that digital media offers a possibility for the user to express their opinion regarding the news. It is not necessary to rely on the users' goodwill while seeking their feedback.

In this study, we define user reliability as follow:

Definition 3.4.1 (Reputation)**.** *A user's likelihood of disseminating true news, i.e., more reliable users share more true news than fake news; less reliable users share more fake news than genuine news.*

3.4.3. User Reputation Calculation Model

The evaluation of trust and reputation is a topic that has received a lot of attention in a variety of fields. Wireless Sensor Networks [**83, 96, 109**], Online Social Networks [**3**], and Quality of Service [**97**]. However, we didn't find a lot of effort dedicated to reputation in the fake news field. Most of them, such as in [**73**], address unreliable user detection as a classification challenge using deep learning methods. In this work, we intend to utilize a probabilistic trust evaluation: Beta Trust Model.

3.4.3.1. Beta Trust Model.

In this part, we define the fundamental concept and mathematically formulate the problem. We apply the same concept as in [**51**]. The objective is to create probabilistic models for user behaviour based on the outcomes of prior experiences. We can predict the likelihood of specific outcomes of the following action using these models.

Assuming that the outcomes are either success s (share real news) or failure f (share fake news), the goal is to represent a user's behaviour using a beta probability distribution across alternative outcomes, either success s or failure f. Bayesian approaches can be used to estimate the parameters of this beta distribution given a sequence of outcomes $h = o_1 \dots o_n$.

With the beta trust model, the outcomes are binary. As a result, we concentrate on the single probability θ_r that a specific user's activity will be successful (reliable). A sequence of n outcomes $h = o_1 \dots o_n$ is a sequence of Bernoulli trials under the assumption of fixed θ_r, and the number of successful outcomes in h is probabilistically distributed according to a binomial distribution.

$$P(\ h \text{ consists of } k \text{ successes}) = \begin{bmatrix} n \\ k \end{bmatrix} \theta_r^k (1-\theta_r)^{n-k}$$

the beta probability density function (pdf) indexed by the parameters α and β

$$f(\theta_r \mid \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)+\Gamma(\beta)}\theta_r^{\alpha-1}(1-\theta_r)^{\beta-1}$$

where Γ is the gamma function, is a conjugate prior to the binomial distribution. That is,if $f(\theta_r \mid \alpha_{pr}, \beta_{pr})$ sen as the a priori pdf of θ_r, then, given a sequence h of outcomes, the resulting a posteriori pdf of θ_r is $f(\theta_r \mid \alpha_{post}, \beta_{post})$ the beta pdf with parameters α_{post} and β_{post} where the a posteriori parameters are related to the a priori ones and the outcome sequence h by the following equations: $\alpha_{post} = N_s(h) + \alpha_{pr}$ and $\beta_{post} = N_f(h) + \beta_{pr}$, where $N_s(h)$ $N_f(h)$ denote the numbers of successful (share real news) and unsuccessful interactions (share fake news)in h, respectively.

Here the estimate for θ_r the probability of having successful interaction, is naturally evaluated as the expected value of θ_r according to its a posteriori pdf. Using the properties of the beta pdf, this expected value is given by:

$$\mathbb{E}[\theta_r] = \frac{\alpha_{post}}{\alpha_{post}+\beta_{post}}$$

A uniform pdf, which assigns equal likelihood to all values of θ_r in the range [0,1], can be represented exactly by a beta distribution with chosen parameters $\alpha = 1$ and $\beta = 1$. Taking the uniform pdf as the a priori pdf for θ_r which just indicates an "unbiased" prior belief about θ_r, as no value is more likely than another,then the parameters of the a posteriori pdf, namely, α_{post} and β_{post}, are related to the sequence h of outcomes as follows:

$$\alpha_{post} = N_s(h) + 1$$

$$\beta_{post} = N_f(h) + 1$$

and the beta estimate for $\theta_r$is therefore given by:

$$\mathbb{E}[\theta_r] = \frac{N_s(h)+1}{N_s(h)+N_f(h)+2}$$

Therefore, the reputation of a user i $\mathbb{R}_i$ is calculated based on the number of shared news, $N_s(h)$ for real news and $N_f(h)$ fake news, by the user i:

$$\boxed{\mathbb{R}_i = \frac{N_s(h)+1}{N_s(h)+N_f(h)+2}} \tag{3.4.1}$$

In this work, we define three types of users based on their reliability: reliable, less reliable and unreliable user. Hence, we need to define thresholds.

According to the equation 3.4.1, when a user reads the same number of fake news as real news, i.e, $N_s(h) = N_f(h)$, his reputation is equal to $\mathbb{R}_i = 0.5$. As a result, he is regarded as

a less reliable user. A specific instance is when a new user enter the system, i.e $N_s(h) = 0$ and $N_f(h) = 0$, his reputation is also equal to $\mathbb{R}_i = 0.5$. Therefore, we define each user i, who have a reputation equal to $\mathbb{R}_i = 0.5$ is a less reliable user. When a user posts more real news than fake news, he is labeled a reliable user, i.e, $\mathbb{R}_i > 0.5$. Hence, when $\mathbb{R}_i < 0.5$, the user is considered as unreliable user.

3.4.3.2. Update Function.

The user status in the system varies dynamically. A well-behaved user may be corrupted at some point due to their news consumption. As a result, trust values must alter dynamically in order to accurately reflect the state of the users. To update the trust value, we adopt the sliding time window approach. The time frame is divided into time slots, and each time slot represents an update cycle. The system will calculate the users' trust value in each time slot, which may be written as: $\mathbb{R}_i(t)$ to denote the reputation value of the user i where $t = 1,2 \dots n$, n denotes the number of time slots. The updated trust value can be expressed as:

$$\boxed{\mathbb{R}_i(t+1)_{update} = \gamma_t \mathbb{R}_i(t) + \gamma_{t+1} \mathbb{R}_i(t+1)}$$

where γ_t and γ_{t+1} represent the weight of the historical trust value and current trust value respectively.

The historical trust values indicate the trust value of previous users. The current trust value is the most recent trust value of users. However, the recent trust value is more significant and has a higher weight. We define, as in [**96**], aging factor θ to describe the damping of trust value, and $\gamma_t = \theta$, $\gamma_{t+1} = 1 - \theta$. In practise, we apply the Simulation parameters $\theta = 0.1$.

Conclusion

We started this chapter by providing a general overview of the second component, FANAR and the problem formulation. Then, we presented the model architecture in this work covering each part: News Modeling and User Modeling, in great depth. Furthermore, we described how we defined and calculated user reliability. The following chapter will be devoted to presenting fake news detection and awareness components.

Chapter 4

Fake news Detection and Awareness

Introduction

In this chapter, we will discuss the two components related to fake news context. First, the **EXMULF**, **EX**plainable **MU**ltimodal Content-based **F**ake News Detection System, is discussed. This component will be used to detect fake news and provides interpretable explanations to users. First, we'll go over the overall architecture. Then we go through each component in detail. Next, we present **FNASY**, a **F**ake **N**ews **A**wareness **SY**stem. We emphasize the significance of awareness in the fake news area. The component's functionality is then described in detail.

4.1. An Explainable Multimodal Content-based Fake News Detection System

4.1.1. Overview

In this section, we present the proposed system for an explainable multimodal content-based fake news detection, named as **EXMULF** (**EX**plainable **MUL**timodal **F**ake news detection). This component, as depicted in the figure 4.1 below, receives a news recommendations as inputs, classifies them (real/fake), and gives explanations using explainable artificial intelligence.

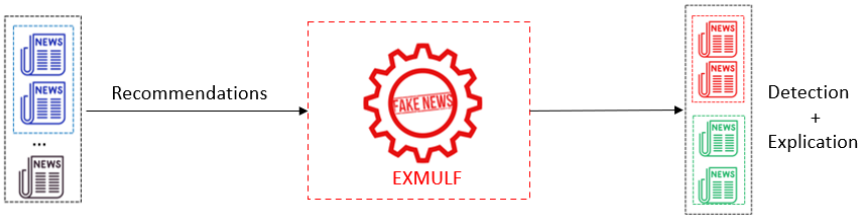


Fig. 4.1. EXMULF Overview.

This chapter introduces a content-based fake news detection system that contains three automated processes to address: 1) multimodal topic modelling, 2) multimodal content-based detection, and 3) multimodal explainable detection. With this in mind, the main contributions of this component are then to:

- Analyze multimodal data within the news content.
- Elaborate a multimodal topic modelling analysis based on the Latent Dirichlet Allocation (LDA) topic model to measure the topic similarity between the text and the image within the online news content.
- Use multimodal data to detect fake news based on Vision-and-Language BERT (VilBERT).
- Generate appropriate multimodal explanations based on Local Interpretable Model-agnostic Explanations (LIME).
- Implement and evaluate our system using two publicly available multimodal datasets (i.e. Twitter and Weibo).

4.1.2. The general architecture

Figure 4.2 illustrates the general architecture of EXMULF. It consists of three major components:

(1) A topic modelling component;
(2) A multimodal content-based fake news detection component (multimodal detector);
(3) And a multimodal explainable detection component (multimodal explainer).

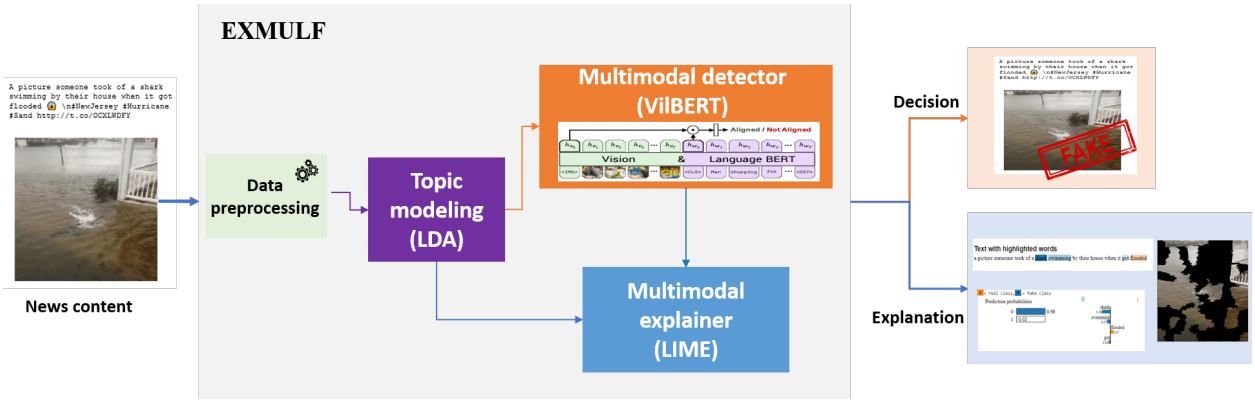


Fig. 4.2. The general architecture of EXMULF.

Figure 4.3 illustrates an overview of the adopted methodology. Specifically, the news content is first provided as input into our system. The text available in the associated image (when applicable) is extracted. Both texts available in the news content and in the

associated image are processed for text analysis. The associated image is also processed for image analysis.

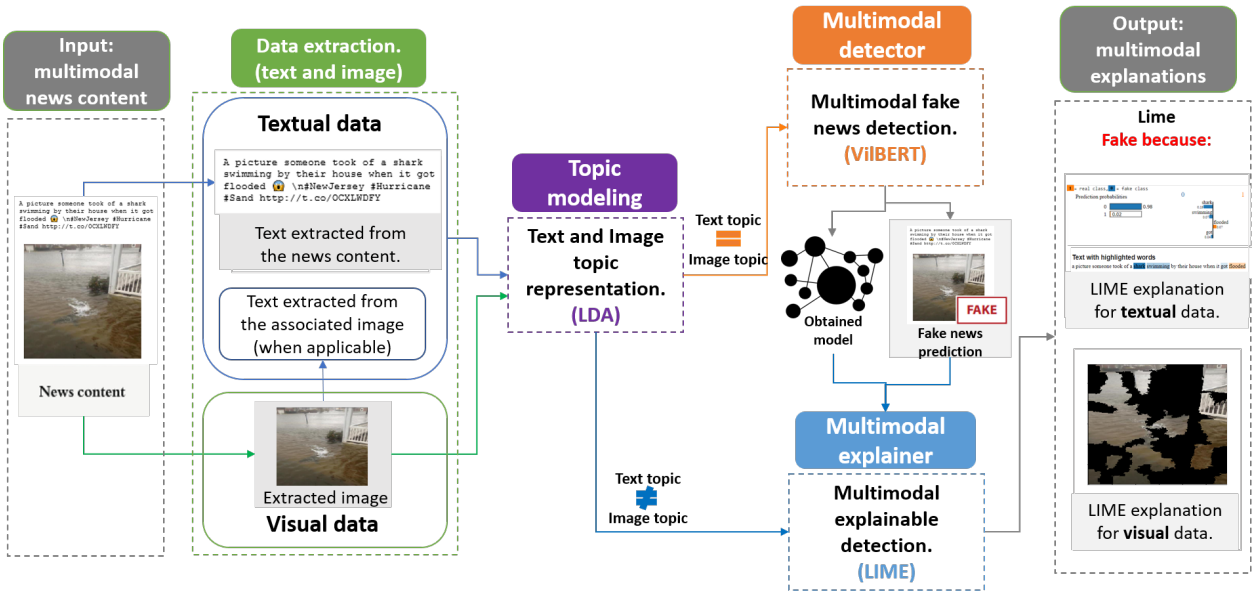


Fig. 4.3. EXMULF methodology overview.

Then the obtained multimodal data (i.e. text and image) are passed to the topic modelling component for topic similarity detection to measure the similarity between both text and image topics. If the captured topics are different, then the news is classified as fake and an explanation based on this will be provided by the multimodal explainer component. Otherwise, the multimodal data obtained will be passed to the multimodal detector component to predict the news veracity based on analyzing the latent task-agnostic joint representations of the text and the associated image. These results are then processed by the multimodal detector component to predict the veracity of the news content. Finally, a decision is rendered, the prediction model as well as the extracted text and image are processed by the multimodal explainer component to generate relevant interpretable explanations to provide to the end-users.

4.1.3. Topic Modeling

The topic representation models consist of topic modelling of both text and image within the online news. Using such an approach is motivated by the fact that the inconsistency (incoherence) between text and image topics in online news can be a major sign that the news is fake. Therefore, the goal of topic modeling is to understand the differences or similarities of topics between the image and the text of the news. On the basis of which a decision on the nature of the input news can be made and explained to OSN users.

The LDA topic Modeling component, is based on the use of the Latent Dirichlet Allocation (LDA) [**10**] which is a probabilistic modelling approach. This method makes it possible

to create topic representations of texts in a corpus by identifying latent semantic structures in the text. Figure 4.4 presents an illustration of the LDA input/output workflow.

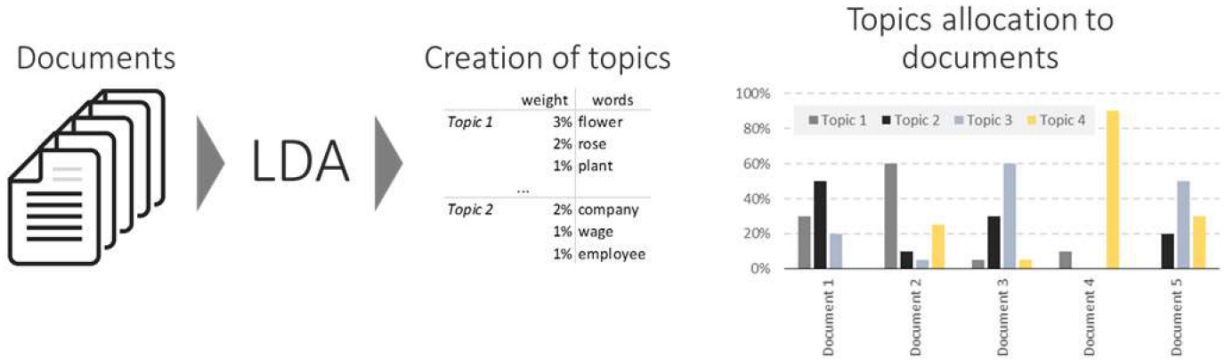


Fig. 4.4. Popular picture used in literature to explain LDA.

By identifying latent semantic patterns in the text, this approach enables the generation of topic representations of texts in a corpus. LDA is a topic model that may be used to categorize text in a document. Using Dirichlet distributions, it builds a topic per document model and a word per topic model. The LDA method is applied to each text document in the collection, providing a list of keywords. Documents are then grouped together to determine the recurring keywords in document groupings. As a result, these clusters of recurring keywords are recognized as a topic shared by multiple papers in the collection.

In this component, we use LDA topic modelling to capture both text and image topics of a given piece of online news. If an inconsistency between the captured topics is found, then our system infers that the news text and its associated image are not aligned. Therefore, the news must have been manipulated and it is then classified as fake.

4.1.4. The Multimodal Detector

The multimodal content-based detection models serves to detect the veracity of a given piece of news based on the multimodal data (text and image) available in its news content (i.e. the text body of the news and its associated image). Using such an approach is motivated by the fact that the news content is the main entity in the deception process, and it is a straightforward and fully available factor in the early stages. As such it can be analyzed and used for the early detection of fake news. However, relying on skeptical auxiliary information captured from the social context of the online news (i.e. social engagement, user response, propagation patterns, etc.) rather than focusing on the news content is not ideal for fake news early detection.

The multimodal content-based detector component, is based on VilBERT (Vision-and-Language BERT) [**53**] which is a model for learning task-agnostic joint representations of

image content and natural language. ViLBERT's architecture is illustrated in figure 4.5 below. The model is made up of two parallel streams for visual (green) and linguistic (purple) processing that interact via novel co-attentional transformer layers. This structure allows for varying depths for each modality and sparse interaction via co-attention. Repeated layers are denoted by dashed boxes with multiplier subscripts.

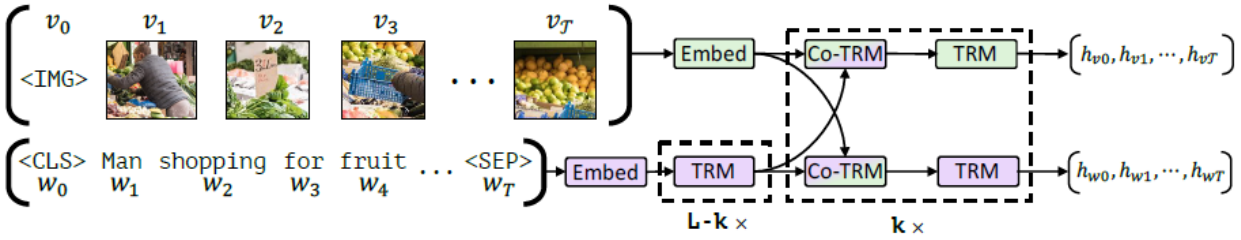


Fig. 4.5. ViLBERT model [**53**].

Images and text inputs are processed separately at first. Text is encoded independently of image features using various transformer layers. The image features are embedded in a form that can be fed into a Transformer; bounding boxes are utilized to locate and select image areas; and a vector is used to store the spatial location of each encoded image section. Then, co-attentional transformer layers are introduced, which employ co-attention to learn the mapping between words in the text input and areas in the image. The model creates a hidden representation that may be used to begin a variety of multimodal tasks.

VilBERT is pretrained on the conceptual captions dataset using two training objectives, masked multimodal learning and image text alignment prediction. The latter is what motivates us the most to employ VilBERT in our multimodal detector component. We chose to use ViLBERT because of its high performance on a variety of visiolinguistic tasks, including visual question answering and image retrieval. However, to apply the pretrained VilBERT model in a multimodal fake news detection/classification task, we fine-tuned it on our datasets in order to acquire visually grounded language understanding in the fake news context. Specifically, we used the multi-task pre-trained model and then we added a linear classification layer of image and text representations to predict whether the news is fake or real.

The multimodal detector component has two major tasks: 1) text processing, and 2) image processing. The image and text are processed in two separate streams. Each stream consists of transformer blocks based on BERT [**22**] and co-attentive layers, which facilitate the interaction between visual and textual modalities. In each co-attentive transformer layer, multi-head attention is computed in the same way as it would be for a standard transformer block, except that the visual modality takes care of the textual modality and vice versa.

In the text processing task, text tokens are generated from the BERT's tokenizer. In the image processing task, images are preprocessed in order to generate regional representations,

including bounding boxes and regional features which are generated with a pre-trained object detection model (MaskRCNN [**37**] in our case, unlike in the original VilBERT model [**53**] where the authors used Faster R-CNN to extract region features). It also encodes the spatial location of the regions. Regional image features and location features are then projected to the same dimension and summed to form the image embedding.

4.1.5. The Multimodal Explainer

Explainable models consist of providing meaningful explanations that aim to let users build trust in the outcome so that they make use of the proposed systems [**70**]. Thus, these explanations help OSN users understand the decision made by our system and "why" a piece of given news news is classified as fake. Consequently, it makes them aware of the danger of such content and influences their future behavior. For instance, a user who is convinced by the explanations provided as to why a piece of given news is fake is unlikely to participate in its dissemination, support or recreation online.

Artificial intelligence applications require trust to aid in decision-making. Otherwise, their advice may be ignored due to a lack of trust. Specifically, if users do not trust a model or prediction, they won't use it [**70**]. In fact, end-users will always prefer solutions that are easy to interpret and understand. Therefore, Explainable AI methods, such as LIME, are helpful to understand how these models use complex mathematical decisions in order to get the corresponding predictions.

Ribeiro et al. [**70**] present LIME as an algorithm that can explain the predictions of any classifier or regressor in a faithful way, by approximating it locally with an interpretable model. Despite the fact that many machine learning models are black boxes, understanding the reasoning behind the model's predictions will undoubtedly help users determine when to trust or not trust its predictions. Figure 4.6 shows an example of a model predicting that a particular patient has the flu. The forecast is then explained by an "explainer", that highlights the symptoms that are most essential to the model. With this knowledge of the model's logic, the doctor may decide whether or not to trust it.

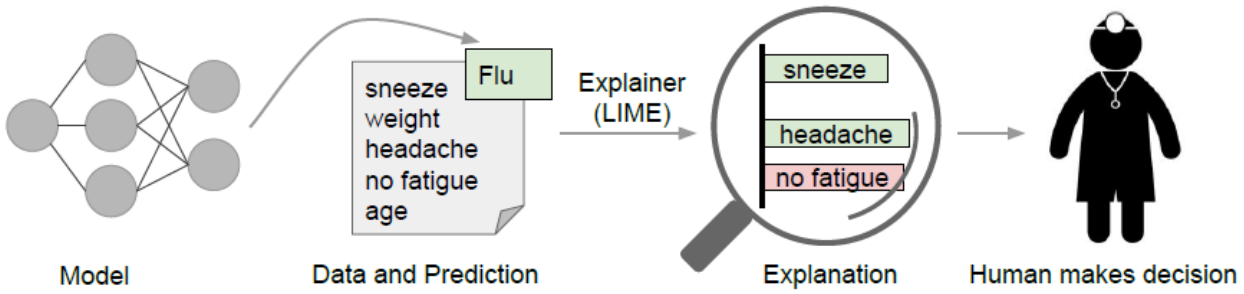


Fig. 4.6. The process of explaining individual predictions [**70**].

The greatest assets of LIME are its accessibility and simplicity. LIME is model agnostic, which means that it can be used with any machine learning model. It provides explanations for almost any given model by treating it as a separate "black-box". In addition, LIME gives local explanations, which are explanations for each observation instead of just the model itself. Furthermore, LIME is interpretable, it offers explanations based on the input features instead of abstract features.

In our system, we use LIME to highlight the features, in both text and image input, that can help classify the news as fake or real. To do so, and after getting the prediction of the multimodal detector, we analyze the text and image separately.

4.2. Fake News Awareness System

4.2.1. The importance of Awareness

As is well established in the physics of complex systems and economics, macro-level events emerge from patterns of individual micro-level behaviour. Minor changes in individual behaviours, when adopted in large numbers, as a result of incentives, restraints, or persuasive ideas, result in macro-level changes in society that can cause benefits or damages at the collective level. The Covid-19 pandemic is a noteworthy example because it has immediately focused public attention towards the influence of individual behaviours on the safety, health, and prosperity of the community. Similarly, if the user receives nudges regarding fake news, they will not consume or spread it. As a result, the spread of fake news will be limited.

Misinformation is a multifaceted issue, with human beings being one of those issues. They are, in fact, the weakest link, due to a lack of awareness. According to recent statistics[1], the number of unintended fake news spreaders on social media is five times larger than the rate of purposeful spreaders. Furthermore, the research reveals that the number of individuals who are confident in their capacity to distinguish reality from fiction is 10 times greater than the percentage of people who are not sure about the truthfulness of what they are sharing. This is a complicated topic since many individuals trust practically anything they read on the Internet, and those who are new to the digital world or have little knowledge, may be easily deceived [**25**].

Through personalized education and persuasion, rewards and incentives, monitoring, tracking, and policing of human behaviours, the advancement of digital technology has enabled support of behavioral changes on a personal level. However, the vast majority of awareness studies emphasize the need of being aware of fake news rather than offering ways to enhance individual awareness. In our method, we provide a personalized awareness component to guide the user to the proper news. This component will notify the user whether

[1]https://www.statista.com/statistics/657111/fake-news-sharing-online/

the news is fake or real and will provide explanations based on prior news that he has clicked on. The awareness component will be presented in the next section.

4.2.2. FNASY Process

In this section, we will go through the system's awareness element. After providing the appropriate recommendations for the user, we must make people aware of fake news. FNASY is designed for this purpose. Figure 4.7 illustrates the general architecture of FNASY.

This module receives the user profile, specifically user reliability, and the EXMULF output for the candidate news, i.e. the label and the related explanation of the news, as inputs. The latter then notifies the user and gives personalized nudges based on user reliability. Specifically, we suggest three levels of awareness: **Simple Awareness**, **Medium Awareness**, and **High Awareness**. This is an essential part of personalization since various users have varying levels of trustworthiness.

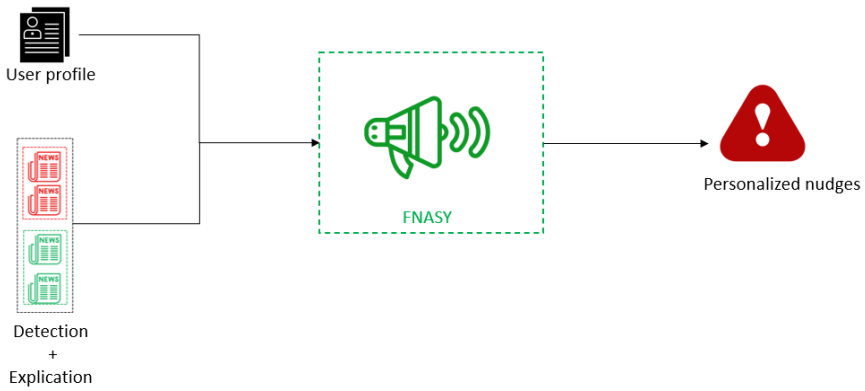


Fig. 4.7. FNASY workflow.

The pseudo-code of the fake news awareness system algorithm is shown in Algorithm 1.

The type of awareness is determined by the user's reliability, as explained in the algorithm. Specifically, if the user reputation value is between 0 and 0.5, indicating that the user is trustworthy and consumes/shares real news, the system will raise limited awareness by simply generating an alert that indicates that the news is fake. In the other case, if the user reputation value is equal to 0.5, which means that the user consumes/shares real news at the same ratio as fake news, the system will provide medium awareness by generating an alert that mention that the news is fake and adding explanations, i.e., why the news is classified as fake. Finally, if the user's reputation is above 0.5, the user is more inclined to interact with fake news rather than real news. In this instance, the system will raise a high level of awareness. In addition to the alerts and explanations, the user will be unable to click on the news, which will be shown in greyscale.

Algorithm 1 FNASY algorithm

Input: UR: User Reliability.
News: Candidate news.
Label: ▷ EXMULF returns the label of the candidate news.
Explication: ▷ EXMULF returns the explications of the candidate news.
Output: awareness
Initialization:
Alert = "*This news is fake*".
Threshold = 0.5
Begin
if Label == "real" **then**
pass ▷ There is no need for awareness if the item is considered as reliable news.
else if Label == "FAKE" **then**
if UR > **Threshold then**
awareness = **Alert** ▷ Simple Awareness.
else if UR = **Threshold then**
awareness = **Alert** + **Explication** ▷ Medium Awareness.
else if UR < **Threshold then**
awareness = **Alert** + **Explication**+ unclickable news ▷ High Awareness.
return **awareness** ▷ return the corresponding awareness
End

Conclusion

We focused on the overall overview of our EXMULF component throughout this chapter. We went over each component's function in detail. We also introduce our final component, FNASY. We went over the algorithm in order to provide a clear vision of the mechanism behind it. The following part of the thesis will be devoted to presenting the experiments.

Part III

Experiments

Chapter 5

FNEWR: A New Dataset

Introduction

Datasets power AI models in the same way that gasoline or electricity fuels vehicles. It is critical to have adequate data for artificial intelligence models to be effective. Furthermore, in order to make actual contributions to the system, we require a suitable collection of data. Therefore, we attempted to create **FNEWR** (**F**ake **NEW**s **R**ecommendation), a dataset that would fulfill our needs.

The next sections cover the dataset's requirements, the inquiry performed to identify relevant existing datasets, the methods used to construct the dataset used in this work, and the strengths and limitations of the resulting data.

5.1. Dataset Challenges

It is quite difficult to generate a dataset that will allow for the exploration of how a proposed system could perform. The dataset should contain information on news, users, user-item interactions revealing which people interacted with which items of news, and labels indicating whether the news is fake or real. It is vital to highlight that we consider explicit user behaviours, such as posting or sharing data for the purposes of our work. Furthermore, the system requires multimodal data, which means that the news content should be represented by both text and image.

To sum up, the dataset should satisfy the following criteria:

- Data on news, users, and user-item interactions;
- News labels;
- Multimodal data.

With this purpose in mind, we investigate, in the next section, some of the most common datasets used for recommender systems in general, and explain why these datasets do not fully meet our needs.

5.2. Existing Datasets Investigation

Most of the existing research in the domain of news recommendation is performed on proprietary datasets, with the exception of only a few publicly available datasets [**92**]. The great majority of the proprietary datasets were retrieved from Google News [**19**], Yahoo News [**60**], Bing News [**50**], and MSN News [**91**].

The available datasets are listed in Table 5.1 below with a presentation of some of their individual characteristics.

Table 5.1. Comparisons of the public datasets for news recommendation.

Dataset	Language	# Users	# News	News Information
Plista[**7**]	German	Unknown	70,353	title, body
Adressa [**34**]	Norwegian	3,083,438	48,486	title, body, category, entities
Globo[**20**]	Portuguese	314,000	46,000	word embeddings of texts
YOW [1]	Unknown	25	383	document ID
Yahoo![2]	English	Unknown	14,180	anonymized word IDs
MIND [**94**]	English	1,000,000	161,013	title, abstract, body, category, entities

These datasets, however, have certain limitations. First, most of the datasets are in languages other than English, making them more specific for certain populations. Second, the YOW dataset, for example, is quite limited in size. As a result, it cannot produce good results and cannot be a suitable alternative for deep learning approaches. Another drawback is that the data collection time is extremely short (at most 14 days in the case of Yahoo), making research on personalization based on long-term user models unfeasible. Furthermore, because the data was gathered during just one specified period, some events that occurred precisely during said period may have resulted in certain biases in the data (Plista). Furthermore, the news items in these datasets are represented by their features, where the actual text of the news stories is anonymized with no further information being given (Yahoo) or with no information at all (YOW). Therefore, making recommendations may be difficult considering the lack of information about the articles. Finally, none of the datasets provided include multimodal information. In addition, as pointed out previously, the few attempts in news recommendations incorporating multimodal content have exploited private datasets, as shown in the table 5.1. Aside from these constraints, all of the datasets are missing labels (real/fake) for the news.

During our assessment, we also examined fake news datasets. The fake news field does not have this issue. Even though there is not yet any acknowledged benchmark dataset for fake news detection due to the difficulties of establishing a clear definition of fake news and the

[1]https://users.soe.ucsc.edu/ yiz/papers/data/YOWStudy/
[2]https://webscope.sandbox.yahoo.com/catalog.php?datatype=l

complexity of gathering appropriate data for analysis [**14**]. However, certain publicly available resources are noteworthy. In [**24**], the authors give a review of twenty-seven common datasets for fake news detection, providing insights into each dataset's properties while comparing them. Unfortunately, these datasets often solely contain news-related information. There are no user-news interactions or information about users.

We identified a study that had the same kind of issues [**27**]. However, the authors aimed to focus on the theme of Corona Virus and did not intend to create a multimodal dataset. Furthermore, their dataset is kept private.

Following this analysis, we discovered that existing datasets in the literature either provide a set of labelled misinforming items without providing information about users or user-item interactions, or provide social media data collections (which contain information about users and items) while providing labels indicating which of those items are misinformation or not providing information about user-item interactions. Furthermore, multimodal news content is tough to obtain. To address all of these problems, we chose to create an adequate dataset in order to build our system and achieve our goals.

5.3. Dataset Creation Methodology

As highlighted in the previous subsection, there are currently no datasets available in the scientific literature that explore how the suggested recommendation system may limit the spread of disinformation. As a result, for the sake of this study and as a first step in this direction, we created our own dataset.

Before initiating the dataset creation process, we must clarify the choices taken at each level, particularly the source of the data. With the emergence of online media, people are increasingly likely to acquire their news online, particularly via social media. Twitter, in particular, has been rapidly consolidating its positioning as a news source. Instead of other channels, many people resort to Twitter for their daily dose of information. According to statistics in [**41**], more than half of Twitter users rely on the platform for news consumption, 48 % of American citizens prefer obtaining their news on social media, and 22 % of journalists use Twitter as their first source of news.

Twitter is a relevant source for data collection. First, by relying on the user's tweets, we can learn about the user's interests and propose personalized news articles that the user would post on Twitter. In fact, several works [**45, 2**] relied on Twitter for the developpment of news recommendation algorithms. Second, many studies [**98, 105, 47, 79, 64**] have focused on social media, notably Twitter, as a source of fake news. Finally, Twitter's APIs makes it possible to collect users, tweets, and interactions. Other platforms (such as Facebook or Whatsapp) only enable data collection from public groups or pages, which implies that user-item interactions access would be restricted and incomplete.

Hence, we decided to work on Twitter data. Furthermore, we needed to collect a set of users and their associated user-item interactions, as repeatedly stated. For that reason, we assumed that whenever a person retweeted (shared) an item (tweet), they had a preference for that item.

Each step as well as the methodology adopted to create this dataset are outlined in figure 5.1 below, and described thoroughly in the upcoming paragraphs.

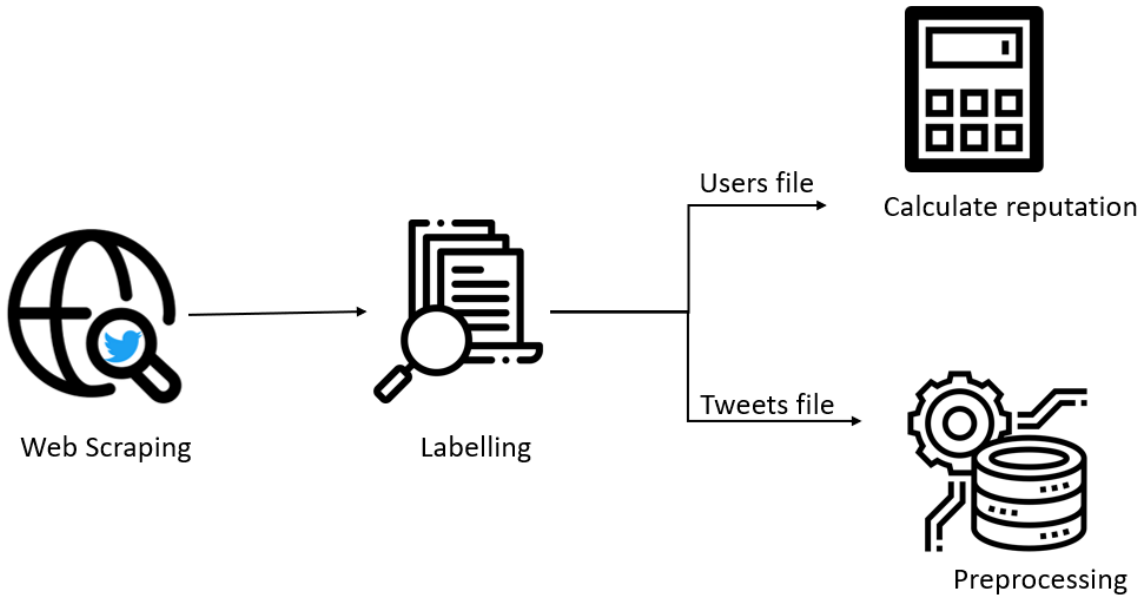


Fig. 5.1. Dataset generation workflow.

Web Scraping: Our starting point is the MediaEval dataset [**11**]. This dataset was retrieved from Twitter and is commonly utilized in the fake news detection field [**80, 105, 47**]. We extracted creators of all tweets in this dataset that were labelled as fake or real. We next chose a random sample of individuals who had retweeted those tweets in order to add more users. In addition, we gathered people that reposted a relevant News handle's tweets (New York Times, Washington Post, and Wall Street Journal) in order to expand the dataset. Once we obtained the user IDs, we used the Twitter API to download their timelines. Tweepy was used to scrape the Twitter.

Because we require a multimodal dataset, we used the image URL returned by Tweepy to download the attached images. To scrape images, we used BeautifulSoup and requests packages since we needed to make a request to retrieve the URL, and then pass the contents of the page through BeautifulSoup so that it can be parsed. Since we needed image data, we used the img tag. After that, the images would be downloaded.

Labeling: We proceed to classify the news after gathering all of the data. We categorize the news using the EXMULF component, see section 4.1, based on its text and image. This is a necessary stage since we require a dataset containing the news label.

Before advancing with the remainder of the stages, we split the dataset into two files:

- A Users file that contains user information such as user id, tweet id for false and true news, and additional information that will be included in the coming phases such as user reputation and user neighbourhood.
- A Tweets file that includes tweet details such as tweet id, tweet text, tweet image, and label.

Calculating reputation: The reputation is computed in accordance with the formula outlined in a previous chapter. This step is for the Users file. We introduce a new column for reputation value. The reputation of a user i $\mathbb{R}_i$ is calculated based on the number of shared news, $N_s(h)$ for real news and $N_f(h)$ fake news, by the user i:

$$\mathbb{R}_i = \frac{N_s(h) + 1}{N_s(h) + N_f(h) + 2}$$

Preprocessing: The preprocessing of the tweets file included the removal of single modality instances, the preprocessing of textual data (i.e. the removal of punctuation, symbols and emoji from the text, as well as URLs) using NLTK library, and the preprocessing of images (i.e. resizing all images to the same equal size) using PIL library.

5.4. Statistics

The FNEWR dataset is presented in this section. After the previously outlined process, the data collection required for the system is ultimately complete.

Table 5.2 below contains some data statistics.

Table 5.2. FNEWR: Statistics.

	#News	#Users
Fake	6922	6463
Real	17017	8760
Total	23939	15413

FNEWR comprised of 23939 collected tweets, 6922 of which are classified fake. It also has 15413 users, of which 6463 only consume fake news. Only 190 users read both fake and true news, whereas 8760 consume only real news.

It's also important to keep in mind that the dataset is separated into two files:

- **Users file:** consist of user id (**users**), the tweets id that he shared (**fake tweets** and **real tweets**) and his reliability (**reputation**).
- **Tweets file:** incorporates the tweet id (**tweetId**), the text within the tweet (**tweetText**), the accompanied image name (**imageId(s)**) and **label** (fake/real).

5.5. The Dataset's Strengths and Weaknesses

As we have demonstrated in the previous sections, collecting datasets that allow us to evaluate how NRS might be improved to limit disinformation dissemination presents considerable obstacles. Our dataset was created with great care and attention to detail. Despite these limitations, it is vital to emphasize that adjusting Recommendation Algorithms (RAs) to prevent misinformation dissemination is a topic that has not previously been addressed despite its importance. To the best of our knowledge, there are no existing datasets or baselines. The created dataset is the first of its type.

Despite the efforts involved in its creation, it is crucial to note that the resulting dataset has certain limitations. To begin with, the dataset is limited in size; it does not include as many items or users as others. Second, the news stories are labelled with EXMULF. Although it worked well and resulted in excellent findings, it was not as effective as the fact-checking methods employed by current fake news databases.

Conclusion

In addition to exploring the existing datasets proposed in the literature for fake news detection, in this chapter we presented the different challenges. Furthermore, we presented the FNEWR dataset creation methodology and some statistics. Thus, in the next chapter, we use FNEWR dataset to assess the ability of the system and we describe in detail the experiments.

Chapter 6

Experiments and Results

Introduction

In this chapter, we will go through the details of the implementation of the methods we described in the previous part. First, we illustrate and describe the test environment and experiment setup, and then we discuss and evaluate the findings.

6.1. Experimental Results and Discussion: EXMULF

The experimental details of the explainable multimodal content-based fake news detection system are provided in this section. We describe the datasets, the tools that were utilized, the findings interpretation, and compare the proposed model to state-of-the-art approaches.

6.1.1. Datasets and Preprocessing

Datasets: We used two publicly available real-world benchmark datasets for our experiments: Twitter[1] and Weibo[2]. Table 6.1 shows the distribution for both datasets after the preprocessing phase.

Table 6.1. Statistics of the datasets used.

Dataset	Train		Test	
	Fake	Real	Fake	Real
Twitter	6841	5009	2564	1217
Weibo	3748	3783	1000	996

The Twitter dataset was released by Boididou et al. [**12**] as a part of Verifying Multimedia Use at MediaEval challenge. This dataset consists of two parts: a training set and a test set.

[1] https://github.com/MKLab-ITI/image-verification-corpus
[2] https://drive.google.com/file/d/14VQ7EWPiFeGzxp3XC2DeEHi-BEisDINn/view

For the Weibo Dataset, Jin et al. [**44**] crawled all the verified false rumor posts from May 2012 to January 2016 on the official rumor debunking system of Weibo (a micro-blogging website in China that encourages users to report suspicious tweets).

Preprocessing : The preprocessing part of Twitter dataset is as following: First, we removed all the instances that only contain text or images because our vision is about multimodal data. Then, for textual data, we removed stop words, punctuation, symbols and emojis. Additionally, the non-English text is translated into English using google translate. The images in the dataset are of differing sizes, so they must be resized before being used in order to match the input size of the neural network. Therefore, for image data, we resized all the images into to be of equal size. Furthermore, we extracted text within the images (when applicable) using the pytesseract library of python (Python-tesseract).

For the Weibo dataset, the preprocessing phase was inspired by the same preprocessing presented by Wang et al. [**89**]. In fact, for image data, we removed duplicate images and odd-sized images to ensure the dataset's integrity. For text data, we proceed the same as for the Twitter dataset while taking the Chinese language into account.

6.1.2. The LDA Topic Modeling

Latent Dirichlet Allocation (LDA) is a topic model that can be used to assign a certain subject to text found in a document. It generates a single topic per document model and a single word per topic model using Dirichlet distributions. Each text document in the collection is subjected to the LDA algorithm, which extracts a list of keywords. Documents are then grouped together in order to understand the recurrent keywords in each groupings of documents. These groups of recurrent keywords are therefore regarded as a topic shared by multiple papers in the collection.

The LDA topic modelling component measures the similarity between text and image topics of the online news. Therefore, in this section, we give experimental settings and results for each task separately.

Topic Modeling for Textual Data.

After preprocessing the text (Tokenization, removal of stop words, lemmatization, and stemming), we create a dictionary containing the number of times a word appears in the training set and compute TF-IDF (term frequency-inverse document frequency) to assess the significance of a term in a document in comparison to a collection or a corpus. In addition to the data and dictionary, we specify the number of topics so as to train the base LDA model. We chose 10 as the number of topics since it gave us the best coherence value.

Different configurations were adopted as shown in Table 6.2. Validation-set refers to our dataset, topics refer to the number of topics (K), the hyperparameter alpha refers to the Document-Topic Density, the hyperparameter beta refers to Word-Topic Density and

Table 6.2. Topic modeling configuration.

Validation-set	Topics	Alpha	Beta	Coherence
74% Corpus	2	0.01	0.01	0.402372
74% Corpus	2	0.01	0.31	0.379257
74% Corpus	2	0.01	0.61	0.378883
74% Corpus	2	0.01	0.91	0.389730
74% Corpus	2	0.01	symmetric	0.379257
....	....	....	....	....
100% Corps	10	asymmetric	0.01	0.491387
100% Corps	10	asymmetric	0.31	0.374487
100% Corps	10	asymmetric	0.61	0.408294
100% Corps	10	asymmetric	0.91	0.317167
100% Corps	10	asymmetric	symmetric	0.451740

coherence refers to the evaluation metric that serves to compare the performance of the model with different hyperparameter settings. We ran these experiments sequentially, one parameter at a time while holding the other constant, and over two separate validation corpus sets (75% Corpus and 100% Corpus). To evaluate the model, we used topic coherence as an intrinsic evaluation metric. Topic Coherence scores a single topic by calculating the degree of semantic similarity between the topic's high-scoring terms. These measures help distinguish between subjects that are semantically interpretable and topics that are statistical inference artifacts.

Topic Modeling for Image Data.

Topic modelling for images presented a unique difficulty since it must interpret both visual and linguistic data, which are two entirely distinct types of data. To achieve this, we employ the Latent Dirichlet Allocation (LDA) method to extract topics from the vocabulary of text data, as well as a pre-trained VGGNet16 model to identify patterns from images. We load the images and the topics and return the samples as a single batch. Then convert the loaded image pixels to Numpy array format using img to array function. Then using Keras Vgg16's preprocess input function, process and prepare the image to load it to the pre-trained VGGNet16 model. Then we train the model to predict themes for the supplied images. To evaluate the model, we load the true topics and the predicted topics and calculate the accuracy. An accuracy of 54% was achieved.

6.1.3. Evaluation Metrics

To evaluate the effectiveness of our classification approach, we employed metrics generally used in Machine Learning and Information Retrieval: Accuracy (Acc), Precision (P), Recall (R) and F1 by class. All these measures were calculated in this thesis using scikit-learn[3],

[3]https://scikit-learn.org

an open machine learning package in Python. In Table 6.3, we give the following confusion matrix to illustrate these metrics in our context.

Table 6.3. Example of confusion matrix.

		Predicted	
		Fake	Real
Actual	Fake	a	c
	Real	b	d

The accuracy (Acc) of a model evaluates how well an approach works in classification. It represents the number of correctly classified data instances over the total number of data instances. Hence, $Acc = \frac{a+d}{a+b+c+d}$ Precision (P) is defined as the probability of true positive samples out of all predicted positive samples (i.e, positive predictive value), $P(fake) = \frac{a}{a+b}$. Recall (R) measures the probability of true positive samples over all the original positive samples, $R(fake) = \frac{a}{a+c}$. Finally, the F1 measure is the harmonic mean between both precision and recall. In this case,$F1(fake) = 2\frac{P(fake)R(fake)}{P(fake)+R(fake)}$.

6.1.4. Multimodal Detector

To evaluate the performance of VilBERT on the fake news detection task, we compared it against other models (baselines), single-modality and multimodal models.

(1) **single-modality models**:

(a) **Text only**: To evaluate the text-based fake news detection model, we fine-tune $BERT_{BASE}$ model. The input of this model is the text, which is fed to the pretrained $BERT_{BASE}$. Also, to determine the importance of the text within the image, we use the model for the text of the news only, $BERT_T$, then the input of the model is the concatenation between the text with news and the text within the news, $BERT_{T+IT}$. This is done to compare the performance of the multimodal models to. Furthermore, it is important to note that for Weibo dataset, we used bert-base-Chinese because it is trained on simplified and traditional Chinese text.

(b) **Image only**: Here, we investigate the images only. For that reason, we use VGG-19, a variant of Visual Geometry Group (VGG) model which in short consists of 19 layers (16 convolution layers, 3 Fully connected layer, 5 MaxPool layers and 1 SoftMax layer) and ResNet-34, a 34 layer residual neural network.

(2) **multimodal models**: For the multimodal model evaluation, we define a fusion model that concatenates $BERT_T$ and ResNet-34 features. Then, a MLP is trained on top of it. After that, we select other existing multimodal models: SpotFake, AMFB, FND-SCTI, HMCAN and BDANN. We chose to compare our results to those of these

models because they are trained on the same datasets that we use (i.e. Twitter and Weibo).

A fair comparison was then made based on four evaluation metrics as presented in Table 6.4, namely the classification accuracy, precision, recall and F1-score metrics stated by the corresponding authors. These evaluation metrics are commonly employed for fake news detection.

The results as shown in Table 6.4 demonstrate that VilBERT outperforms the baseline models described above in terms of accuracy.

Table 6.4. EXMULF Results.

Dataset	Model		Acc	Fake News			Real News		
				P	R	F1	P	R	F1
Twitter	Text only	$BERT_T$	0.572	0.602	0.586	0.597	0.543	0.553	0.544
		$BERT_{T+IT}$	0.577	0.612	0.574	0.598	0.551	0.564	0.556
	Image only	ResNet-34	0.624	0.712	0.567	0.6	0.558	0.72	0.62
		VGG-19	0.596	0.698	0.522	0.593	0.531	0.698	0.597
	Multi-modal	Fusion	0.7695	0.820	0.726	0.779	0.719	0.798	0.748
		SpotFake [**79**]	0.7777	0.751	0.900	0.82	0.832	0.606	0.701
		AMFB [**47**]	0.883	0.89	**0.95**	0.92	**0.87**	0.76	0.741
		HMCAN [**64**]	0.897	**0.971**	0.801	0.878	0.853	**0.979**	**0.912**
		BDANN [**108**]	0.830	0.810	0.630	0.710	0.830	0.930	0.880
		VilBERT	**0.898**	0.934	0.92	**0.926**	0.859	0.88	0.869
Weibo	Text only	$BERT_T$	0.680	0.731	0.715	0.709	0.667	0.676	0.669
		$BERT_{T+IT}$	0.682	0.739	0.72	0.71	0.672	0.684	0.673
	Image only	ResNet-34	0.694	0.701	0.634	0.698	0.698	0.711	0.699
		VGG-19	0.633	0.640	0.635	0.637	0.637	0.641	0.639
	Multi-modal	Fusion	0.8152	0.865	0.734	0.88	0.764	0.889	0.74
		SpotFake [**79**]	0.8923	0.902	**0.964**	0.932	0.847	0.656	0.739
		AMFB [**47**]	0.832	0.82	0.86	0.84	0.85	0.81	0.83
		FND-SCTI [**105**]	0.834	0.863	0.780	0.824	0.815	0.892	0.835
		HMCAN [**64**]	0.885	0.920	0.845	0.881	0.856	0.926	**0.890**
		BDANN [**108**]	0.842	0.830	0.870	0.850	0.850	0.820	0.830
		VilBERT	**0.9204**	**0.946**	0.948	**0.946**	**0.879**	**0.893**	0.885

In this study, each dataset was divided into two parts: 80% was assigned to training and 20% to testing. Although VilBERT was originally designed for various vision-and-language challenges, recent research has indicated that the learning of visiolinguistic feature representations may be transferred across tasks. As a result, we fine-tune ViLBERT across datasets by passing the element-wise product of the final image and text representations into a learned classification layer. Moreover, we referenced the Facebook study GitHub repository [**54**].

6.1.5. Multimodal Explainer

For the explanation part, we used LIME for both image and text. Figure 6.1 is a tweet input example to illustrate how LIME works. This tweet is classified as fake news by the multimodal detector component based on VilBERT.

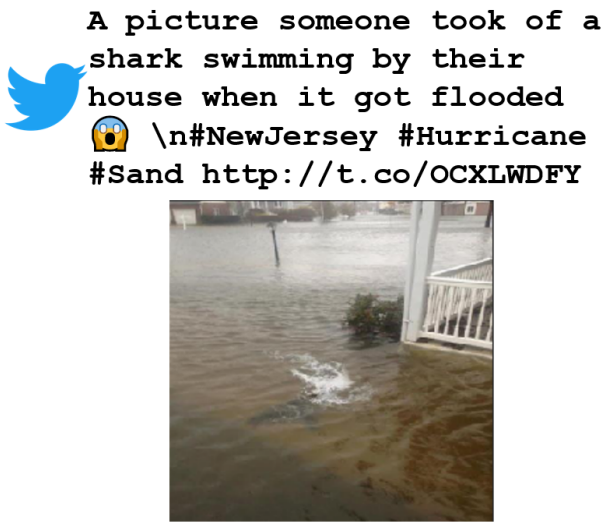


Fig. 6.1. Input tweet example.

For image data, see Figure 6.2, the explanations are created by generating a new dataset of perturbations around the instance to be explained. The output or class of each generated perturbation is predicted with the model. The importance of each perturbation is determined by measuring its distance from the original instance to be explained. These distances are converted to weights by mapping the distances to a zero-one scale using a kernel function. All this information: the new generated dataset, its class predictions and its weights are used to fit a simpler model (linear model), that can be interpreted. The attributes of the linear model, coefficients for the case of a linear model, are then used to generate explanations. For that reason, the first step is to create perturbations of the image. To do so, we turn on and off some of the superpixels using the quickshift segmentation algorithm. In the second step, we used the model to predict the class of the newly generated images. After that, the distance between the original image and each of the perturbed images is computed in order to measure the importance (i.e. weights) of each perturbation. To do that, we choose to use cosine similarity. Finally, a weighted linear regression model is fitted using perturbations, predictions, and weights. Each coefficient in the linear model corresponds to a superpixel in the segmented image. These coefficients represent the importance of each superpixel in predicting the corresponding class. The original study [**70**] provided this strategy.

(a) (b) (c)

Fig. 6.2. (a) presents the original image (b) shows the superpixels that are generated using the quickshift segmentation algorithm (c) shows the area of the image that produced the prediction of the class (fake, in our case)

On the other hand, we use LIME Text Explainer for textual data. For this, we add a separate text instance to the interpreter. In the case of text data, different versions of the original text are created, in which a certain number of different, randomly selected words are removed. This new artificial data is then assigned to different categories (fake/real). Hence, through the presence or absence of certain keywords we can see their influence on the classification of the selected text. The original publication [**70**] offered this method.

The output of LIME is a list of explanations, reflecting the contribution of each feature to the prediction of a data sample. The return value summarizes the contribution of each word to the assignment of the text instance to a specific class (i.e. fake, real). The visualisations, see Figure 6.3, help us understand what words in a text have the greatest influence in terms of the model's final prediction.

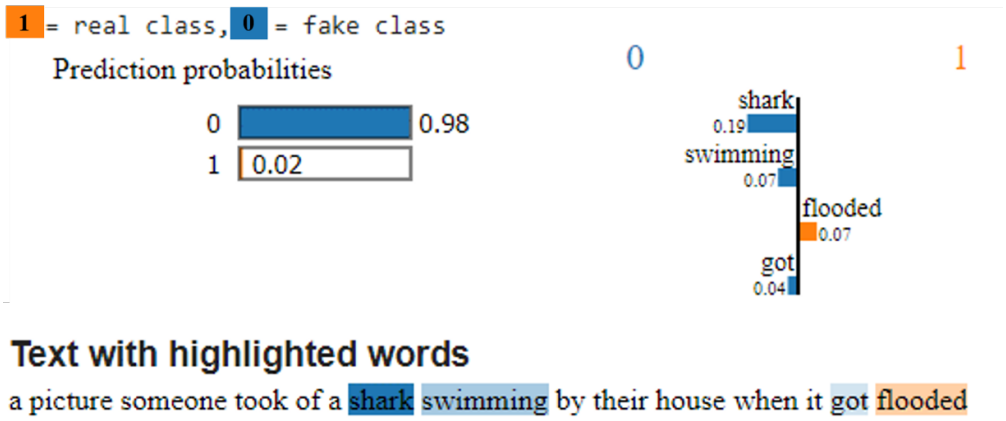


Fig. 6.3. LIME explanations for textual data.

6.1.6. Discussion

To test the topic modelling component, we execute its two sub-models (topic modelling for textual data and topic modelling for image data) on a subset of 1000 samples from a labelled dataset. The goal is to confirm that if the text and image inside the online news have distinct topics, the news is fake. 722 news out of 1000 samples were found with text and images containing diverse themes, knowing that 496 out of 722 samples (i.e. 68 percent) were first identified as fake.

It's also important to note that the detection models perform differently in the two datasets. In fact, they produce better results with the Weibo dataset than with the Twitter dataset. These findings are connected to the fact that most of the images in the Weibo dataset appear to be more involved . Furthermore, because Weibo is a Chinese dataset, the length of certain sentences after segmentation exceeds the length of sentences in the Twitter dataset.

When we investigate the performance of single-modality models, we discover that the image-only model performs worse than the text-only model, indicating that text appears to be considerably more important than visual information in the identification of fake news. Despite the fact that BERT demonstrates qualifying skills in terms of performance evaluation for both single-modality and multimodality, its performance is still insignificant when compared to multimodal approaches that supplement textual characteristics with visual features.

Based on the results, it is plausible to assume that combining image and text is advantageous since it provides better performance than either image or text by themselves. The pre-trained ViLBERT also outperforms the other baselines. This means that learning the semantic relationship between visual and linguistic is transferable across activities. The pre-trained multi-task model performs exceptionally well when it comes to matching image and text signals. However,VilBERT does not always have the best values for recall, precision, and F1 score. It performs better on the Weibo dataset than on the Twitter dataset (imbalanced dataset).

6.2. Experimental Results and Discussion: FANAR

This part comprises the experimental details of the fake news aware news recommendation system. We present the techniques used, the interpretation of the data, and a comparison of the suggested model to state-of-the-art methodologies, as well as its variants.

6.2.1. Experimental Settings

6.2.1.1. Datasets.

To evaluate FANAR approach we used the newly generated dataset named FNEWR, that

is described in the previous chapter .

6.2.1.2. Evaluation Metrics.

There are several metrics that may be used to empirically evaluate the performance of news recommender systems. In this work, three popular metrics are adopted to evaluate FANAR, namely the Area Under Curve (AUC) score, Mean Reciprocal Rank (MRR) and Normalized Discounted Cumulative Gain (nDCG).

$$AUC = \frac{|\{(i,j)|Rank(p_i) < Rank(n_j)\}|}{N_p N_n}$$

$$MRR = \frac{1}{N_p} \sum_{i=1}^{N_p} \frac{1}{Rank(p_i)}$$

$$nDCG@K = \frac{\sum_{i=1}^{K} (2^{r_i} - 1)/log_2(1+i)}{\sum_{i=1}^{N_p} 1/log_2(1+i)}$$

where:

- N_p and N_n are the numbers of positive and negative samples, respectively.
- p_i is the predicted score of the i-th positive sample.
- n_j is the predicted score of the j-th negative sample.
- r_i is a relevance score of news with the i-th rank, which is 1 for clicked news and 0 for non-clicked news.
- K number of the top K recommendation list.

6.2.1.3. Parameter Settings.

We built our proposed method using Pytorch[4], a well-known Python toolkit for neural networks. The dataset was divided into three parts: 60% as a training set for learning parameters, 20% as a validation set for tuning hyperparameters, and 20% as a testing set for final performance comparison. We experimented with the values [8, 16, 32, 64, 128, 256], for the embedding size d, [32, 64, 128, 512] for the batch size and and [0.0005, 0.001, 0.005, 0.01, 0.05, 0.1] for the learning rate. Furthermore, we experimentally set the hidden layer size to be equal to the embedding size and the activation function to be ReLU. We used three hidden layers for all of the neural components. Model parameters were randomly initialized for all neural network approaches. The settings for the baseline algorithms were initialized as described in the appropriate articles and then carefully tuned to obtain optimal performance.

6.2.1.4. Baselines.

We compare the proposed method with many baseline methods, including:

[4]https://pytorch.org/

- **DKN**[**87**][5], learning news representations via a knowledge-aware CNN model.
- **NAML**[**90**][6], use attentive multiview learning to learn news representations.
- **EBNR**[**60**][7], an embedding-based news recommendation approach that uses an autoencoder to learn news embeddings and a GRU network to learn user representations.
- **Wide&Deep**[**17**][8], a neural recommender with a wide linear component and a deep neural network component.

Only news texts are examined in these approaches. We selected these models because their source code is publicly available; unfortunately, many other techniques did not disclose their source code. It is also worth mentioning that, in order to do a fair comparison, we must test all of the algorithms using the same dataset.

6.2.2. Performance Evaluation

We assess the performance of our methodology by comparing it to the previously stated baseline approaches. The AUC, MRR, and nDCG@5 values were reported. Table 6.5 showcases the experimental results.

Table 6.5. Performance comparison of different methods 1.

Model	AUC	MRR	nDCG@5
DKN	60.43	19.95	21.77
NAML	61.63	21.98	23.77
EBNR	60.64	20.54	22.16
Wide&Deep	58.66	19.24	21.13
FANAR	**61.74**	**21.72**	**23.56**

The findings indicate that our FANAR technique, which takes into account visual information in news, outperforms competing approaches based only on textual content, such as DKN and EBNR. This is due to the fact that individuals commonly pick news items based not just because of their interest in news texts, but also on the attractiveness of news images.

As a result, the visual information supplied by news images is extremely beneficial when it comes to learning proper news representations for recommendation purposes. While existing news recommendation approaches neglect image-related information, our FANAR method combines both textual and visual news information into news representation learning while modelling their intrinsic relatedness for enhanced news content comprehension. Thus, FANAR can accomplishes better results for news recommendations.

[5]https://github.com/hwwang55/DKN
[6]https://github.com/wuch15/IJCAI2019-NAML
[7]https://github.com/Leavingseason/rnn$_r ecsys$
[8]https://github.com/kaitolucifer/wide-and-deep-learning-keras

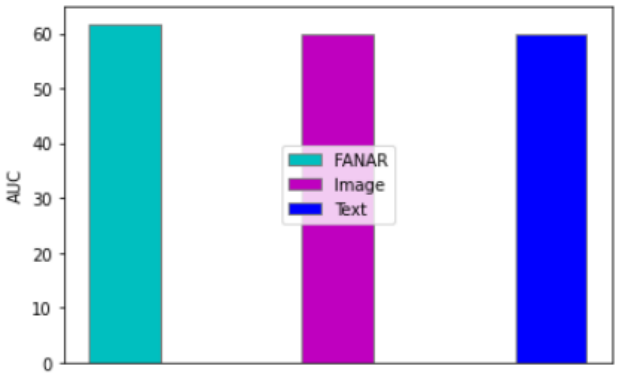


Fig. 6.4. Effectiveness of multimodal information.

6.2.3. Model Analysis

An ablation study is provided in this study. We compare the FANAR model to several variants in order to assess different parts of the model.

6.2.3.1. Effect of multimodal information.

We evaluate the performance of our FANAR technique to two variants that only analyze images or texts in the news representation. Figure 6.4 illustrates the results. According to our findings, both the news text and the image are important for the learning of news representations for recommendation purposes. It demonstrates that both textual and visual news information are effective for the comprehension of news content and predicting consumer interest. Furthermore, including multimodal news information can increase recommendation performance even further, thus demonstrating that incorporating multimodal news information can assist in the learning of appropriate news representations.

6.2.3.2. Effect of eliminating unreliable users.

In this section, we propose comparing the model to a variation FANAR-RFN in which we maintain all neighbors without regards regarding user reliability, or whether the news item is fake or real. Table 6.6 presents the experimental findings.

Table 6.6. Performance comparison of different methods 2.

Model	AUC	MRR	nDCG@5
DKN	60.43	19.95	21.77
NAML	61.63	21.98	23.77
EBNR	60.64	20.54	22.16
Wide&Deep	58.66	19.24	21.13
FANAR	**61.74**	**21.72**	**23.56**
FANAR-RFN	**62.44**	**22.36**	**24.58**

The variant FANAR-RFN outperforms FANAR results. This may be because of the elimination unreliable of neighbors, see section 3.3.2.2. FANAR outperforms some other news recommendation algorithms, as is indicated in Table 6.5 above. However, the comparison is not fair since the FANAR model excludes certain neighbors (those that are deemed to be unreliable). As a result, some data is missing that may improve the outcomes.

6.2.4. Beyond Accuracy Evaluation

The primary aim of the study is to see if FANAR, the recommendation algorithm, can assist to reduce the spread of misleading content. Thus, we separate the efficiency (i.e., accuracy), a dimension commonly linked with the performance evaluation of these algorithms, from the actual measurement of misinformation propagation With this objective in mind, we suggest the **total fake news (TF)** as an assessment metric, which measures the number of misleading items in FANAR's recommendation lists over all the list.

$$TF = \frac{\text{total of recommended fake news}}{\text{total of recommended news}}$$

The evaluation metric findings are reported in the table 6.7 below.

Table 6.7. Beyond Accuracy Evaluation.

Model	TF@5	TF@10
FANAR	0.2	0.3
FANAR-RFN	0.4	0.4

The experiments reveal that FANAR is more successful at reducing fake news recommendations than its counterpart FANAR-RFN, which does not take fake news context into account. In fact, the difference between the two lies in the adaptation found within the model. In order to compensate for the misinformation within the RAs, we reduce the number of neighbors used within the model by eliminating the unreliable ones.

6.2.5. Discussion

To evaluate the FANAR model's performance, we first compare it to several established baselines. The results reveal that it outperforms several baselines that simply analyze textual data. We conduct a model analysis to uncover several variations in order to dig deeper into the performance of the suggested model. First, we examine the influence of multimodal data. The findings suggest that both textual and visual news information are extremely effective for the comprehension of news content and the prediction of user interest. Furthermore, including multimodal news information can increase recommendation performance even further, thus demonstrating that incorporating multimodal news information can assist in the developpment of accurate news representations.

Following this, we compare the FANAR model to the FANAR- RFN. It is worth noting that FANAR removes a lot of information, such as such as those deemed to be untrustworthy neighbors who may have the same interest as the candidate user. According to the results, FANAR-RFN outperforms FANAR. The purpose of this effort was not to outperform existing recommendation algorithms in terms of accuracy but to minimize the propagation of fake news. We modified the algorithm to achieve this goal. As a result, in order to assess the performance of the FANAR model in accomplishing this aim, we must modify the assessment measures in the same way. The results reveal that FANAR successfully reduces the number of fake news items in the recommendation lists.

6.3. Full Scenario

This section presents a detailed scenario to help better understand how the system works. The recommendation data is represented as graph data with two graphs. These two graphs include a neighbors graph denoting the relationships between users, and a user-news graph denoting interactions between users and items.

Assuming that we have the following graph, in Figure 6.5, between users (left side) and news (right side). For example, user U1 has shared/consumed news n5,n6, and n8. Each user is represented by User ID (The anonymous ID of a user), History (The previews shared/viewed news), The neighbors, and User reliability. The news item is characterized by News ID, Text, Image, and Label (fake, red dots/real, blue dots).

As stated before, the reputation is the user's likelihood of disseminating true news. In other words, more reliable users share more true news than fake news, less reliable users share more fake news than genuine news. The user reputation is calculated based on the equation (3.4.1).

For example, user U3 consumed three news, one fake news (n4) and one real news (n3). According to the equation, the user reputation can be calculated as follow:

$$\mathbb{R}_{U3} = \frac{1+1}{1+1+2} = 0.5 = 0.5$$

Hence, U6 is less reliable user.

FNEWR dataset is composed of two files: Tweets files that incorporates the tweet id, text and image of he tweet, and Users file that consists of user id, tweets id that he shared ans his reliability. For the example presented above, we can represent the Users file as in Figure 6.6.

Suppose that U3 is the candidate user. As previously stated, for the U3 neighbourhood only the **most reliable neighbors** are taken into account. The neighbors are:U1,U5,U4, and U7. However, since U4 is unreliable user, he will be eliminated.

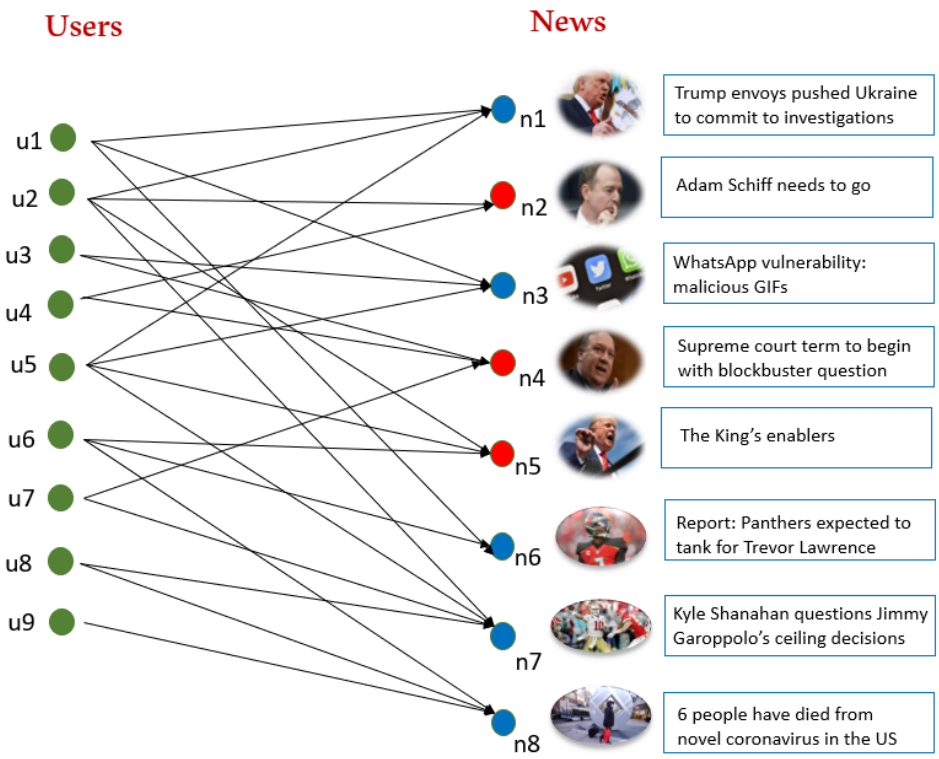


Fig. 6.5. Graph between users and news.

User	History	User reliability	Neighbors
U1	n1,n3,n6	Reliable	U2,U3, U5,U6
U2	n1,n2,n5,n7	Less reliable	U1,U5,U4,U3,U6,U7,U8
U3	n3,n4	Less reliable	U1,U5,U4,U7
U4	n2,n4	unreliable	U2,U3,U7
U5	n1,n3,n5,n7	Reliable	U1,U2,U3,U6,U7,U8
U6	n5,n6,n8	Reliable	U2,U5,U1,U8,U9
U7	n4, n7	Less reliable	U3,U4,U2,U5,U8
U8	n7,n8	Reliable	U2,U5,U7,U6,U9
U9	n8	Reliable	U6,U8

Fig. 6.6. Dataset Example.

FANAR, the personalized fake news aware recommender system, will provide some recommendations to individual users based on their preferences as determined by their profile. FANAR is mainly an adaptation recommendation algorithm that aids in the reduction of fake news spread by avoiding unreliable neighbors.

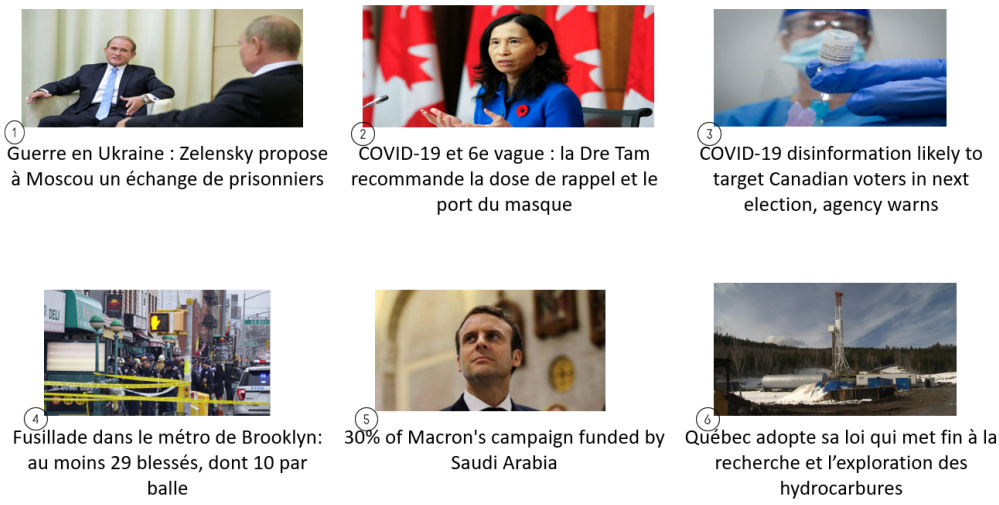


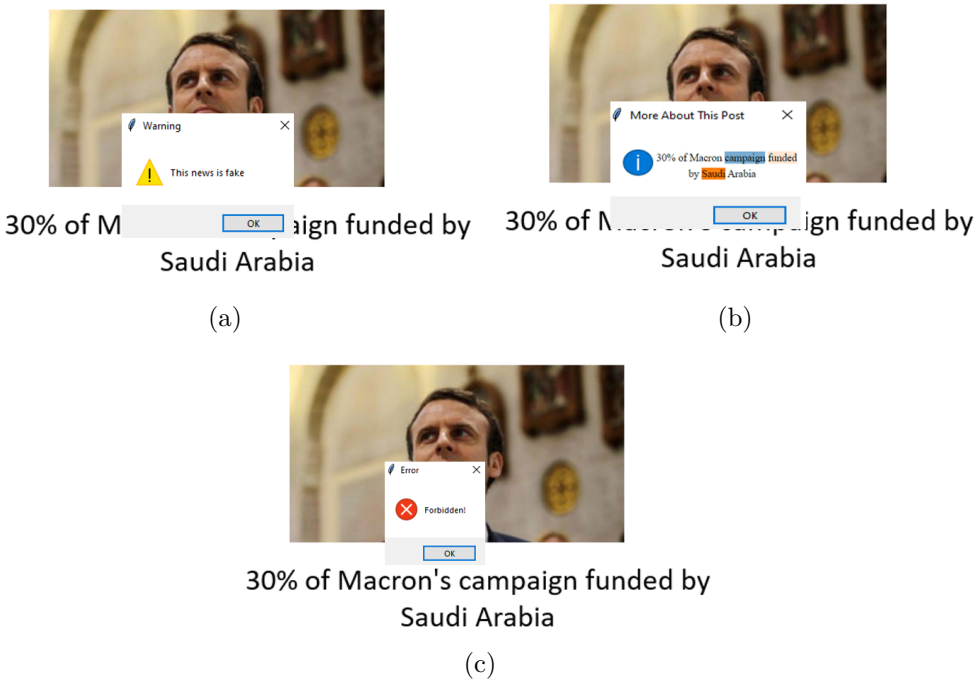
Fig. 6.7. FANAR results for the candidate user.

Assuming that FANAR, for the candidate user, gives the following news, as shown in Figure 6.7. After that, EXMULF, EXplainable MUltimodal Content-based Fake News Detection System, classifies the news based on the text and image. This component receives a news recommendations as inputs, classifies them (real/fake), and gives explanations using explainable artificial intelligence. In this scenario, the third and fifth news are categorised as fake news, while the others are real.

Suppose that candidate user U3 picks news number five, which is a fake news. Then, FNASY, the fake news awareness system notifies the user and gives personalized nudges based on user reliability. As discussed before, FNASY can provide three types of awareness as presented in Figure 6.8.

Conclusion

In this chapter, we explored the performance of current approach for the fake news problem. Particularly, we explored FNEWR, the created dataset. Our results provide an interesting perspective on the current performance of the different components. Furthermore, we provided an overall interpretation of the results.



(a) (b)

(c)

Fig. 6.8. (a) presents the Simple Awareness (b) shows the Medium Awareness (c) shows the High Awareness.

Chapter 7

Conclusion

In this chapter, we summarize the main contributions of this thesis. Furthermore, we also offer a discussion on future research directions that can be explored.

7.1. Summary of Results

In this thesis, we explored the practical utility of automated methods for the detection of fake news in order to reduce its spread on digital platforms. Particularly, (1) we developed a novel news recommendation algorithm. We adapted the algorithm to reduce the number of fake news in the recommendation lists (2) We also developed a system for fake news detection based on multimodal data, i.e, image and textual data. This system does not only classify the news but also provides explanations, and last (3) we proposed a fake news awareness algorithm, capable of providing personalized alerts based on user reliability. To the best of our knowledge, this is the first initiative that integrates detection, mitigation, and awareness in the context of fake news. Such studies are categorized by research goals, which are summarized below:

- **RG1 - Adapting recommendation algorithm to avoid fake news:**
  We developed a methodology for recommending news with the intention of minimizing the spread of fake news. To that end, we propose a probabilistic model named Beta Trust model, to calculate user reputation, and then classify users into two categories : reliable and unreliable users. Following that, we adjust the recommendation algorithm by removing untrustworthy people from the candidate user's neighbourhood. Our goal is to limit the amount of fake news in the recommendation list since, as previously noted, existing recommendation algorithms promote the propagation of fake news. For the recommendation process, We chose to utilize Graph Neural Networks since their main role is to use neural networks to repeatedly extract feature information from local graph neighbors.
  The findings demonstrate that our FANAR method, which considers visual information in news, beats competing algorithms that only consider textual content. In order

to analyze the model, we also compare several variants of it. First, we investigated the impact of multimodal data on recommendations. We discovered that including multimodal news information improves recommendation performance. We also analyzed the effect of removing untrustworthy users. The trials show that FANAR is more effective at lowering fake news recommendations than its variant FANAR-RFN, which does not consider fake news context.

- **RG2 - Exploring the multimodal data available in news content to detect fake news and provide explanations:**
  To accomplish the second research goal, we suggest EXMULF, a multimodal content-based fake news detection system that takes as input the textual and visual information within the content of an online news post (i.e. text and image), detects whether the post is fake or real, and explains to users the reasoning behind system's decisions. We concentrate on news content since it is a critical aspect for early detection since it is completely available in the early stages, as opposed to auxiliary information (i.e. social interaction, user response, propagation patterns, etc.) which can only be gathered after the news has propagated. To predict the alignment between the text and the corresponding image, we employ VilBERT (Vision-and-Language BERT) multimodal alignment, and LIME (Local Interpretable Model-agnostic Explanations) to provide the user with an interpretable explanation. Detailed experiments were carried out on two publicly available multimodal datasets (Twitter and Weibo). The experimental findings reveal that our system outperforms ten current state-of-the-art methods in the detection of fake news. As a result, integrating textual, visual, and text-image topic modelling analysis with multimodal explainability is quite beneficial when it comes to tackling the challenge of fake news detection. To the best of our knowledge, this is the first work to use VilBERT and LIME models to develop a fully explainable multimodal content-based fake news detection method.

- **RG3 - Raising awareness about fake news:**
  We believe that raising user awareness, educating and showing them the reasons why the news is fake will help reduce the spread of fake news. Hence, we include an awareness module in the post-recommendation part. Personalized awareness is provided based on the user profile.

7.2. Future Research Directions

Despite the significance of the findings obtained in this thesis, including contributions from concurrent work, combating fake news is a typical adversarial issue that needs ongoing

research. Thus, this thesis opens a vast swath of questioning that can be further explored in future works.

To improve the performance of the fake news detection method, we propose to include audio and video as multimodal input data. Furthermore, we intend to enhance the visual representations so as to improve the efficacy of explainability supplied to users. More assessments will be conducted in order to improve the performance of the multimodal explanation.

Furthermore, larger data volumes will lead to improved performance and the exploration of new methodologies in the future. We also suggest exploring other recommendation algorithms to see how they affect the spread of fake news. Moreover, we believe it is critical to assess the efficiency of the awareness component by asking users if the notifications help them avoid fake news.

Finally, as a following step, we intend to establish an end-to-end system. For the moment being, each component acts independently. Our goal is to construct a process, as illustrated in Figure 2.3, that conducts the workflow from beginning to end and offers a fully working solution.

7.3. Bibliographical Contributions

The following publications were created as a result of the major findings of this thesis:

- Accepted Paper in FPS, The 14th International Symposium on Foundations Practice of Security 2021, EXMULF: An Explainable Multimodal Content-based Fake News Detection System, Sabrine Amri, Dorsaf Sallami and Esma Aïmeur.(To appear)
- Submitted paper for the 16th ACM Conference on Recommender Systems (RecSys 2022).

References

[1] Collins 2017 word of the year shortlist, 2017. Last accessed 05 March 2022.

[2] Fabian Abel, Qi Gao, Geert-Jan Houben et Ke Tao : Twitter-based user modeling for news recommendations. *In Twenty-Third International Joint Conference on Artificial Intelligence*. Citeseer, 2013.

[3] Sami Al-Yazidi, Jawad Berri, Muhammad Al-Qurishi et Majed Al-Alrubaian : Measuring reputation and influence in online social networks: A systematic literature review. *IEEE Access*, 8: 105824–105851, 2020.

[4] Raed Alharbi, Minh N Vu et My T Thai : Evaluating fake news detection models from explainable machine learning perspectives. *In ICC 2021-IEEE International Conference on Communications*, pages 1–6. IEEE, 2021.

[5] Mingxiao An, Fangzhao Wu, Chuhan Wu, Kun Zhang, Zheng Liu et Xing Xie : Neural news recommendation with long-and short-term user representations. *In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 336–345, 2019.

[6] Oberiri Destiny Apuke et Bahiyah Omar : Fake news proliferation in nigeria: Consequences, motivations, and prevention through awareness strategies. *Humanities and Social Sciences Reviews*, 8(2):318–327, 2020.

[7] Jon Atle Gulla, Kevin C Almeroth, Mozhgan Tavakolifard et Frank Hopfgartner : *Workshop and challenge on news recommender systems.* ACM, 2013.

[8] Monika Bednarek et Helen Caple : 'value added': Language, image and news values. *Discourse, context & media*, 1(2-3):103–113, 2012.

[9] Bimal Bhattarai, Ole-Christoffer Granmo et Lei Jiao : Explainable tsetlin machine framework for fake news detection with credibility score assessment. *arXiv preprint arXiv:2105.09114*, 2021.

[10] David M Blei, Andrew Y Ng et Michael I Jordan : Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.

[11] Christina Boididou, Katerina Andreadou, Symeon Papadopoulos, Duc-Tien Dang-Nguyen, Giulia Boato, Michael Riegler, Yiannis Kompatsiaris *et al.* : Verifying multimedia use at mediaeval 2015. *MediaEval*, 3(3):7, 2015.

[12] Christina Boididou, Symeon Papadopoulos, Markos Zampoglou, Lazaros Apostolidis, Olga Papadopoulou et Yiannis Kompatsiaris : Detection and visualization of misleading content on twitter. *International Journal of Multimedia Information Retrieval*, 7(1):71–86, 2018.

[13] Charles F Bond Jr et Bella M DePaulo : Accuracy of deception judgments. *Personality and social psychology Review*, 10(3):214–234, 2006.

[14] Alessandro Bondielli et Francesco Marcelloni : A survey on fake news and rumour detection techniques. *Information Sciences*, 497:38–55, 2019.

[15] Joanna M Burkhardt : History of fake news. *Library Technology Reports*, 53(8):5–9, 2017.

[16] Chong Chen, Min Zhang, Yiqun Liu et Shaoping Ma : Neural attentional rating regression with review-level explanations. *In Proceedings of the 2018 World Wide Web Conference*, pages 1583–1592, 2018.

[17] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir *et al.* : Wide & deep learning for recommender systems. *In Proceedings of the 1st workshop on deep learning for recommender systems*, pages 7–10, 2016.

[18] Evandro Cunha, Gabriel Magno, Josemar Caetano, Douglas Teixeira et Virgilio Almeida : Fake news as we feel it: perception and conceptualization of the term "fake news" in the media. *In International Conference on Social Informatics*, pages 151–166. Springer, 2018.

[19] Abhinandan S Das, Mayur Datar, Ashutosh Garg et Shyam Rajaram : Google news personalization: scalable online collaborative filtering. *In Proceedings of the 16th international conference on World Wide Web*, pages 271–280, 2007.

[20] Gabriel de Souza Pereira Moreira, Felipe Ferreira et Adilson Marques da Cunha : News session-based recommendations using deep neural networks. *In Proceedings of the 3rd Workshop on Deep Learning for Recommender Systems*, pages 15–23, 2018.

[21] Ronald Denaux et Jose Manuel Gomez-Perez : Linked credibility reviews for explainable misinformation detection. *In International Semantic Web Conference*, pages 147–163. Springer, 2020.

[22] Jacob Devlin, Ming-Wei Chang, Kenton Lee et Kristina Toutanova : Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[23] Riccardo Di Massa, Maurizio Montagnuolo et Alberto Messina : Implicit news recommendation based on user interest models and multimodal content analysis. *In Proceedings of the 3rd international workshop on Automated information extraction in media production*, pages 33–38, 2010.

[24] Arianna D'Ulizia, Maria Chiara Caschera, Fernando Ferri et Patrizia Grifoni : Fake news detection: a survey of evaluation datasets. *PeerJ Computer Science*, 7:e518, 2021.

[25] Stephanie Edgerly, Rachel R Mourão, Esther Thorson et Samuel M Tham : When do audiences verify? how perceptions about message and source influence audience verification of news headlines. *Journalism & Mass Communication Quarterly*, 97(1):52–71, 2020.

[26] Miriam Fernández et Alejandro Bellogín : Recommender systems and misinformation: The problem or the solution? *In OHARS@RecSys*, 2020.

[27] Miriam Fernández, Alejandro Bellogín et Iván Cantador : Analysing the effect of recommendation algorithms on the amplification of misinformation. *arXiv preprint arXiv:2103.14748*, 2021.

[28] Paulo Márcio Souza Freire, Flávio Roberto Matias da Silva et Ronaldo Ribeiro Goldschmidt : Fake news detection based on explicit and implicit signals of a hybrid crowd: An approach inspired in meta-learning. *Expert Systems with Applications*, 183:115414, 2021.

[29] Marco Furini, Silvia Mirri, Manuela Montangero et Catia Prandi : Untangling between fake-news and truth in social media to understand the covid-19 coronavirus. *In 2020 IEEE Symposium on Computers and Communications (ISCC)*, pages 1–6. IEEE, 2020.

[30] Suyu Ge, Chuhan Wu, Fangzhao Wu, Tao Qi et Yongfeng Huang : Graph enhanced representation learning for news recommendation. *In Proceedings of The Web Conference 2020*, pages 2863–2869, 2020.

[31] Anastasia Giachanou, Guobiao Zhang et Paolo Rosso : Multimodal fake news detection with textual, visual and semantic information. *In International Conference on Text, Speech, and Dialogue*, pages 30–38. Springer, 2020.

[32] Anastasia Giachanou, Guobiao Zhang et Paolo Rosso : Multimodal multi-image fake news detection. *In 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, pages 647–654. IEEE, 2020.

[33] Alex Graves : Long short-term memory. *Supervised sequence labelling with recurrent neural networks*, pages 37–45, 2012.

[34] Jon Atle Gulla, Lemei Zhang, Peng Liu, Özlem Özgöbek et Xiaomeng Su : The adressa dataset for news recommendation. *In Proceedings of the international conference on web intelligence*, pages 1042–1048, 2017.

[35] Taha Hassan : Trust and trustworthiness in social recommender systems. *In Companion Proceedings of The 2019 World Wide Web Conference*, pages 529–532, 2019.

[36] Sarah Hawa, Lanita Lobo, Unnati Dogra et Vijaya Kamble : Combating misinformation dissemination through verification and content driven recommendation. *In 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, pages 917–924. IEEE, 2021.

[37] Kaiming He, Georgia Gkioxari, Piotr Dollár et Ross Girshick : Mask r-cnn. *In Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[38] Linmei Hu, Chen Li, Chuan Shi, Cheng Yang et Chao Shao : Graph neural news recommendation with long-term and short-term interest modeling. *Information Processing & Management*, 57(2):102142, 2020.

[39] Linmei Hu, Siyong Xu, Chen Li, Cheng Yang, Chuan Shi, Nan Duan, Xing Xie et Ming Zhou : Graph neural news recommendation with unsupervised preference disentanglement. *In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4255–4264, Online, juillet 2020. Association for Computational Linguistics.

[40] Bei Hui, Lizong Zhang, Xue Zhou, Xiao Wen et Yuhui Nian : Personalized recommendation system based on knowledge embedding and historical behavior. *Applied Intelligence*, pages 1–13, 2021.

[41] Allan Jay : 85 twitter statistics you must know: 2021/2022 market share analysis data. Last accessed 20 February 2022.

[42] Zhenyan Ji, Mengdan Wu, Jirui Liu et José Enrique Armendáriz Íñigo : Attention-based graph neural network for news recommendation. *In 2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2021.

[43] Zhenyan Ji, Mengdan Wu, Hong Yang et José Enrique Armendáriz Íñigo : Temporal sensitive heterogeneous graph neural network for news recommendation. *Future Generation Computer Systems*, 2021.

[44] Zhiwei Jin, Juan Cao, Han Guo, Yongdong Zhang et Jiebo Luo : Multimodal fusion with recurrent neural networks for rumor detection on microblogs. *In Proceedings of the 25th ACM international conference on Multimedia*, pages 795–816, 2017.

[45] Nirmal Jonnalagedda et Susan Gauch : Personalized news recommendation using twitter. *In 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, volume 3, pages 21–25. IEEE, 2013.

[46] Jan Kirchner et Christian Reuter : Countering fake news: A comparison of possible solutions regarding user acceptance and effectiveness. *Proceedings of the ACM on Human-computer Interaction*, 4(CSCW2):1–27, 2020.

[47] Rina Kumari et Asif Ekbal : Amfb: Attention based multimodal factorized bilinear pooling for multimodal fake news detection. *Expert Systems with Applications*, 184:115412, 2021.

[48] Lukas Kurasinski et Radu-Casian Mihailescu : Towards machine learning explainability in text classification for fake news detection. *In 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 775–781. IEEE, 2020.

[49] Xiaohan Li, Mengqi Zhang, Shu Wu, Zheng Liu, Liang Wang et S Yu Philip : Dynamic graph collaborative filtering. *In 2020 IEEE International Conference on Data Mining (ICDM)*, pages 322–331. IEEE, 2020.

[50] Jianxun Lian, Fuzheng Zhang, Xing Xie et Guangzhong Sun : Towards better representation learning for personalized news recommendation: a multi-channel deep fusion approach. *In IJCAI*, pages 3805–3811, 2018.

[51] Bin Liu et Geng Yang : Probabilistic trust evaluation with inaccurate reputation reports. *International Journal of Distributed Sensor Networks*, 11(6):736286, 2015.

[52] Kuan-Chieh Lo, Shih-Chieh Dai, Aiping Xiong, Jing Jiang et Lun-Wei Ku : All the wiser: Fake news intervention using user reading preferences. *In Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, pages 1069–1072, 2021.

[53] Jiasen Lu, Dhruv Batra, Devi Parikh et Stefan Lee : Vilbert: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks. *arXiv preprint arXiv:1908.02265*, 2019.

[54] Jiasen Lu, Vedanuj Goswami, Marcus Rohrbach, Devi Parikh et Stefan Lee : 12-in-1: Multi-task vision and language representation learning. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10437–10446, 2020.

[55] Yi-Ju Lu et Cheng-Te Li : Gcan: Graph-aware co-attention networks for explainable fake news detection on social media. *arXiv preprint arXiv:2004.11648*, 2020.

[56] Gabriel Machado Lunardi, Guilherme Medeiros Machado, Vinicius Maran et José Palazzo M de Oliveira : A metric for filter bubble measurement in recommender algorithms considering the news domain. *Applied Soft Computing*, 97:106771, 2020.

[57] Deepak Mangal et Dilip Kumar Sharma : Fake news detection with integration of embedded text cues and image features. *In 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*, pages 68–72. IEEE, 2020.

[58] Priyanka Meel et Dinesh Kumar Vishwakarma : Han, image captioning, and forensics ensemble multimodal fake news detection. *Information Sciences*, 567:23–41, 2021.

[59] Sina Mohseni, Fan Yang, Shiva Pentyala, Mengnan Du, Yi Liu, Nic Lupfer, Xia Hu, Shuiwang Ji et Eric Ragan : Machine learning explanations to prevent overtrust in fake news detection. *arXiv preprint arXiv:2007.12358*, 2020.

[60] Shumpei Okura, Yukihiro Tagami, Shingo Ono et Akira Tajima : Embedding-based news recommendation for millions of users. *In Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1933–1942, 2017.

[61] Yitong Pang, Jianing Tong, Yiming Zhang et Zhihua Wei : Dacnn: Dynamic attentive convolution neural network for news recommendation. *In Proceedings of the 2020 5th International Conference on Mathematics and Artificial Intelligence*, pages 161–166, 2020.

[62] Anish Patankar, Joy Bose et Harshit Khanna : A bias aware news recommendation system. *In 2019 IEEE 13th International Conference on Semantic Computing (ICSC)*, pages 232–238. IEEE, 2019.

[63] Piotr Przybyła et Axel J Soto : When classification accuracy is not enough: Explaining news credibility assessment. *Information Processing & Management*, 58(5):102653, 2021.

[64] Shengsheng Qian, Jinguang Wang, Jun Hu, Quan Fang et Changsheng Xu : Hierarchical multi-modal contextual attention network for fake news detection. *In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 153–162, 2021.

[65] Yongye Qian, Pengpeng Zhao, Zhixu Li, Junhua Fang, Lei Zhao, Victor S Sheng et Zhiming Cui : Interaction graph neural network for news recommendation. *In International Conference on Web Information Systems Engineering*, pages 599–614. Springer, 2020.

[66] Shaina Raza et Chen Ding : News recommender system: A review of recent progress, challenges, and opportunities. *arXiv preprint arXiv:2009.04964*, 2020.

[67] Shaina Raza et Chen Ding : Deep dynamic neural network to trade-off between accuracy and diversity in a news recommender system. *arXiv preprint arXiv:2103.08458*, 2021.

[68] Shaina Raza et Chen Ding : News recommender system: a review of recent progress, challenges, and opportunities. *Artificial Intelligence Review*, pages 1–52, 2021.

[69] Julio CS Reis, André Correia, Fabricio Murai, Adriano Veloso et Fabrício Benevenuto : Explainable machine learning for fake news detection. *In Proceedings of the 10th ACM conference on web science*, pages 17–26, 2019.

[70] Marco Tulio Ribeiro, Sameer Singh et Carlos Guestrin : " why should i trust you?" explaining the predictions of any classifier. *In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144, 2016.

[71] Victoria L Rubin, Yimin Chen et Nadia K Conroy : Deception detection for news: three types of fakes. *Proceedings of the Association for Information Science and Technology*, 52(1):1–4, 2015.

[72] Giancarlo Ruffo, Alfonso Semeraro, Anastasia Giachanou et Paolo Rosso : Surveying the research on fake news in social media: a tale of networks and language. *arXiv e-prints*, pages arXiv–2109, 2021.

[73] Giuseppe Sansonetti, Fabio Gasparetti, Giuseppe D'aniello et Alessandro Micarelli : Unreliable users detection in social media: Deep learning techniques for automatic detection. *IEEE Access*, 8:213154–213167, 2020.

[74] TYSS Santosh, Avirup Saha et Niloy Ganguly : Mvl: Multi-view learning for news recommendation. *In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1873–1876, 2020.

[75] Priyanshi Shah et Ziad Kobti : Multimodal fake news detection using a cultural algorithm with situational and normative knowledge. *In 2020 IEEE Congress on Evolutionary Computation (CEC)*, pages 1–7. IEEE, 2020.

[76] Kai Shu, Limeng Cui, Suhang Wang, Dongwon Lee et Huan Liu : defend: Explainable fake news detection. *In Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 395–405, 2019.

[77] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang et Huan Liu : Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter*, 19(1):22–36, 2017.

[78] Amila Silva, Yi Han, Ling Luo, Shanika Karunasekera et Christopher Leckie : Propagation2vec: Embedding partial propagation networks for explainable fake news early detection. *Information Processing & Management*, 58(5):102618, 2021.

[79] Shivangi Singhal, Rajiv Ratn Shah, Tanmoy Chakraborty, Ponnurangam Kumaraguru et Shin'ichi Satoh : Spotfake: A multi-modal framework for fake news detection. *In 2019 IEEE fifth international conference on multimedia big data (BigMM)*, pages 39–47. IEEE, 2019.

[80] Chenguang Song, Nianwen Ning, Yunlei Zhang et Bin Wu : A multimodal fake news detection model based on crossmodal attention residual and multichannel convolutional neural networks. *Information Processing & Management*, 58(1):102437, 2021.

[81] Hao Tan et Mohit Bansal : Lxmert: Learning cross-modality encoder representations from transformers. *arXiv preprint arXiv:1908.07490*, 2019.

[82] Dhananjay Theckedath et RR Sedamkar : Detecting affect states using vgg16, resnet50 and se-resnet50 networks. *SN Computer Science*, 1(2):1–7, 2020.

[83] V Umarani, K Soma Sundaram et D Jayashree : Enhanced beta trust model in wireless sensor networks. *In 2016 International conference on information communication and embedded systems (ICICES)*, pages 1–5. IEEE, 2016.

[84] Dinesh Kumar Vishwakarma, Deepika Varshney et Ashima Yadav : Detection and veracity analysis of fake news via scrapping and authenticating the web search. *Cognitive Systems Research*, 58:217–229, 2019.

[85] Nguyen Vo et Kyumin Lee : The rise of guardians: Fact-checking url recommendation to combat fake news. *In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 275–284, 2018.

[86] Heyuan Wang, Fangzhao Wu, Zheng Liu et Xing Xie : Fine-grained interest matching for neural news recommendation. *In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 836–845, 2020.

[87] Hongwei Wang, Fuzheng Zhang, Xing Xie et Minyi Guo : Dkn: Deep knowledge-aware network for news recommendation. *In Proceedings of the 2018 world wide web conference*, pages 1835–1844, 2018.

[88] Limin Wang, Sheng Guo, Weilin Huang et Yu Qiao : Places205-vggnet models for scene recognition. *arXiv preprint arXiv:1508.01667*, 2015.

[89] Yaqing Wang, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Kishlay Jha, Lu Su et Jing Gao : Eann: Event adversarial neural networks for multi-modal fake news detection. *In Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining*, pages 849–857, 2018.

[90] Chuhan Wu, Fangzhao Wu, Mingxiao An, Jianqiang Huang, Yongfeng Huang et Xing Xie : Neural news recommendation with attentive multi-view learning. *arXiv preprint arXiv:1907.05576*, 2019.

[91] Chuhan Wu, Fangzhao Wu, Mingxiao An, Jianqiang Huang, Yongfeng Huang et Xing Xie : Npa: neural news recommendation with personalized attention. *In Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 2576–2584, 2019.

[92] Chuhan Wu, Fangzhao Wu, Yongfeng Huang et Xing Xie : Personalized news recommendation: A survey. *arXiv preprint arXiv:2106.08934*, 2021.

[93] Chuhan Wu, Fangzhao Wu, Tao Qi et Yongfeng Huang : Mm-rec: Multimodal news recommendation. *arXiv preprint arXiv:2104.07407*, 2021.

[94] Fangzhao Wu, Ying Qiao, Jiun-Hung Chen, Chuhan Wu, Tao Qi, Jianxun Lian, Danyang Liu, Xing Xie, Jianfeng Gao et Winnie Wu : Mind: A large-scale dataset for news recommendation. *In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3597–3606, 2020.

[95] Shiwen Wu, Fei Sun, Wentao Zhang et Bin Cui : Graph neural networks in recommender systems: a survey. *arXiv preprint arXiv:2011.02260*, 2020.

[96] Xiaoling Wu, Junjie Huang, Jie Ling et Lei Shu : Bltm: beta and lqi based trust model for wireless sensor networks. *IEEE Access*, 7:43679–43690, 2019.

[97] Jianlong Xu, Yindong Chen et Changsheng Zhu : A qos-based user reputation measurement method for web services. *Communication Engineering*, page 470, 2018.

[98] Junxiao Xue, Yabo Wang, Yichen Tian, Yafei Li, Lei Shi et Lin Wei : Detecting fake news by exploring the consistency of multimodal data. *Information Processing & Management*, 58(5):102610, 2021.

[99] Fan Yang, Shiva K Pentyala, Sina Mohseni, Mengnan Du, Hao Yuan, Rhema Linder, Eric D Ragan, Shuiwang Ji et Xia Hu : Xfake: explainable fake news detector with visualizations. *In The World Wide Web Conference*, pages 3600–3604, 2019.

[100] Kehua Yang, Shaosong Long, Wei Zhang, Jiqing Yao et Jing Liu : Personalized news recommendation based on the text and image integration. *CMC-COMPUTERS MATERIALS & CONTINUA*, 64(1):557–570, 2020.

[101] Shuo Yang, Kai Shu, Suhang Wang, Renjie Gu, Fan Wu et Huan Liu : Unsupervised fake news detection on social media: A generative approach. *In Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 5644–5651, 2019.

[102] Di You, Nguyen Vo, Kyumin Lee et Qiang Liu : Attributed multi-relational attention network for fact-checking url recommendation. *In Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 1471–1480, 2019.

[103] Boyang Yu, Jiejing Shao, Quan Cheng, Hang Yu, Guangli Li et Shuai Lü : Multi-source news recommender system based on convolutional neural networks. *In Proceedings of the 3rd International Conference on Intelligent Information Processing*, pages 17–23, 2018.

[104] Hua Yuan, Jie Zheng, Qiongwei Ye, Yu Qian et Yan Zhang : Improving fake news detection with domain-adversarial and graph-attention neural network. *Decision Support Systems*, page 113633, 2021.

[105] Jiangfeng Zeng, Yin Zhang et Xiao Ma : Fake news detection for epidemic emergencies via deep correlations between text and images. *Sustainable Cities and Society*, 66:102652, 2021.

[106] Lemei Zhang, Peng Liu et Jon Atle Gulla : Dynamic attention-integrated neural network for session-based news recommendation. *Machine Learning*, 108(10):1851–1875, 2019.

[107] Mengqi Zhang, Shu Wu, Xueli Yu, Qiang Liu et Liang Wang : Dynamic graph neural networks for sequential recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 2022.

[108] Tong Zhang, Di Wang, Huanhuan Chen, Zhiwei Zeng, Wei Guo, Chunyan Miao et Lizhen Cui : Bdann: Bert-based domain adaptation neural network for multi-modal fake news detection. *In 2020 international joint conference on neural networks (IJCNN)*, pages 1–8. IEEE, 2020.

[109] Jin Zhao, Jifeng Huang et Naixue Xiong : An effective exponential-based trust and reputation evaluation system in wireless sensor networks. *IEEE Access*, 7:33859–33869, 2019.

[110] Xinyi Zhou, Jindi Wu et Reza Zafarani : Safe: Similarity-aware multi-modal fake news detection. *Advances in Knowledge Discovery and Data Mining*, 12085:354, 2020.

[111] Xinyi Zhou et Reza Zafarani : Fake news: A survey of research, detection methods, and opportunities. *arXiv preprint arXiv:1812.00315*, 2, 2018.

[112] Qiannan Zhu, Xiaofei Zhou, Zeliang Song, Jianlong Tan et Li Guo : Dan: Deep attention neural network for news recommendation. *In Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 5973–5980, 2019.