Misleading Connection Mining: Scopes, Computational Challenges and Future Directions

Anonymous Authors

No Institute Given

Abstract. With open and easy access to the internet, people all around the world consume and publish information through online platforms. The invention of social media has added an extra dimension to this information ecosystem. The problem with information being all around us is that sometimes the quality of information can be compromised, leading to information disorder. In this paper, we are particularly interested in one type of information disorder, *Misleading Connection*, which is defined as a scenario *"When headlines, visuals or captions do not support the content."* We focus on providing a comprehensive overview of *Misleading Connection*, including formal defining it, and characterizing it from multiple perspectives. We also identify some challenges of the *Misleading Connection* detection task, suggest some possible solution techniques and discuss some future research directions.

Keywords: Misleading Connection · Social Media · Data Mining.

1 Introduction

We are surrounded by an enormous amount of information generated from different online sources. People have numerous ways to obtain information about current affairs from a variety of channels, blogs, discussion forums, etc. The advent of social media has changed the pattern of our information sharing and consumption. As people spend a considerable amount of time interacting online through social media platforms, there is a proliferation of rapidly disseminated content. From independently acting individuals to media organizations, a wide variety of entities are creating and publishing content over social media. The types of published content are as varied as the types of publishers, and include opinions, news articles, images, and much more. Not only is online content being created on a massive scale, social media sharing allows it be be rapidly and exponentially disseminated. Now that information can be created and disseminated at speeds and scales unparalleled in human history, earnest endeavors that rely on widespread access to "good information" have the potential for rapid, sweeping impact.

But, as has been the case with many technological advances over the centuries, the social media ecosystem has a dark side as well: sometimes the information flowing through it is bad. We will use the term information disorder to refer to these cases where compromised information is being spread, and look at two categories in of information disorder in particular: Misinformation and Disinformation. Misinformation refers to the case where the information is false, but the person who is disseminating it believes that it is true. With Disinformation, not only is the information itself compromised, but so too are the intentions of the person who is disseminating; that is, Disinformation is the knowing and malicious spreading of lies. [1]. Wardle at al. [52] identified 7 types of mis- and disinformation. In this paper, we highlight a specific problem within this realm, *Misleading Connection* . *Misleading Connection* is defined as a scenario "*When headlines, visuals or captions do not support the content*. Let's look at an example of a news article published on usatoday.com-

Example 1. **Headline**: Starbucks will close 8,000 US stores May 29 for racial-bias training [2]
Content: Starbucks plans to close more than 8,000 U.S. stores for several hours next month to conduct racial-bias training for nearly 175,000 workers. This comes after two black men were arrested in one of its stores in Philadelphia. (April 18)

Just after reading the headline, an average reader may think that the Starbucks will be permanently shutting down 8,000 of their stores. But in the content we find that it was planning to close for several hours. So, here the headline is misrepresenting the content and the initial impression created from the headline went wrong after going through the full content. The way the headline is misleading the content, we term this as *Misleading Connection* .

In different platforms such as social media, news media, online review sites two pieces of information can be falsely connected. Post messages, visuals, captions etc. are responsible to create initial impression on social media sites and sometimes the impression is not backed up by the full/whole content. When people scroll through their feeds on social media account without clicking through on articles (which often happens), misleading connections can be especially deceptive. Most of the time people dont read the full articles rather they skim the headlines before being drawn to whatever draws their attention. Gabielkov et al.[19] reported that 59% of the shared links on Twitter were never clicked before sharing it, suggesting that in many cases people only read the headlines. Another study by Ecker et al.[12] shows that even after reading the article in full, a reader is likely to be left with their initial impression gained from the headline. So, *Misleading Connection* can lead to an environment of misinformation and hence we need a solution to fight against this malpractice. Although the problem has much negative impact on the information ecosystem, it is relatively an unexplored area. Chesney et al. [8, 55] studied the incongruency between headline and news body which is a type of *Misleading Connection* . But, we argue that *Misleading Connection* is more than the inconruency problem. There are many other ways a *Misleading Connection* can be made for misleading people such

[1] https://en.unesco.org/fightfakenews
[2] https://www.usatoday.com/videos/news/nation/2018/04/18/starbucks-close-8000-us-stores-may-29-racial-bias-training/33943675/

as using manipulated images, impersonating, showing misleading search result, etc. In this paper, we present an overview of *Misleading Connection* , define it formally, analyze its characteristics, and formulate its detection task. We also investigate the associated feature representations, identify the challenges and outline future research directions.

2 Overview of Misleading Connection

2.1 Misleading Connection as an Information Disorder

The extensive and misuse of the phrase "Fake News" makes it unable to explain the scale of information disinformation and the handbook published by UNESCO [24] recommends using misinformation and disinformation instead. Wardle et al. have identified seven types of mis- and disinformation and the False connection is one of them [53]. It is defined as a scenario *"When headlines, visuals or captions do not support the content"*. While this definition holds the concept from a high level, we argue that a formal and more concrete definition is necessary to identify the associated challenges and possible solutions. So, we choose to use the term "Misleading connection" instead of False connection while they both intent the same meaning but cover different aspects.

Misleading connection is a different problem from other information disorders, for example, rumours [17], false news [6], hoaxes [1], etc. Whether its sole purpose is to mislead people by creating the wrong impression for the unseen content, the problem is different from traditional fake news problem where the contents of the articles are verifiably false. In example 1, the headline is misleading the content by withholding important information (stores will keep close for several hours), but the content is not entirely false (e.g. they didn't take any decision to keep close the store) to make it fake news.

Many may think the problem we are discussing here is a clickbait [37] problem. But we argue that the problem lies in a different spectrum than clickbait. Clickbait headline tricks readers into clicking the link, but a misleading connection may not force the reader to click but still can misguide the readers. In the example 1, people might not find it attractive enough to click, but by withholding important information, it is misleading the readers.

So, the unique characteristics associated with the misleading connection give it a separate identity and urge for comprehensive reviews from different perspectives.

2.2 Misleading Connection: Definition

Although the idea of *Misleading Connection* has been introduced in recent literature, the term is yet to be defined and formally characterized. In this section, we take an attempt to formally define the term *Misleading Connection* with respect to the current internet ecosystem, characterize and categorize it from multiple perspectives.

First, we introduce and define some necessary terminologies and notations. After that, we formally define *Misleading Connection* with the help of these notations.

Definition 1. *Element: A typical HTML object such as hyperlink, image, video, text, button, etc. We denote an element as e.*

Definition 2. *Action: An event (e.g., click, tap) that a user can do to traverse from one set of elements to another set of elements. We denote it by α.*

Definition 3. *Access Content: An access content is a set of elements such that if a user conducts an action α over it, it gives the user access to another set of elements that was out of sight of the user before α was initiated. We denote an access content with $\mathcal{A}$. If an access content has n access elements, we write $\mathcal{A} = \{e_1, e_2, ..., e_n\}$.*

Definition 4. *Hidden Content: It is the set of elements that an $\mathcal{A}$ gives access to upon an action α. We denote a hidden content with $\mathcal{H}$ and write, $\mathcal{A}\vec{\alpha}\mathcal{H}$. If a hidden content has n access elements, we write $\mathcal{H} = \{e_1, e_2, ..., e_n\}$.*

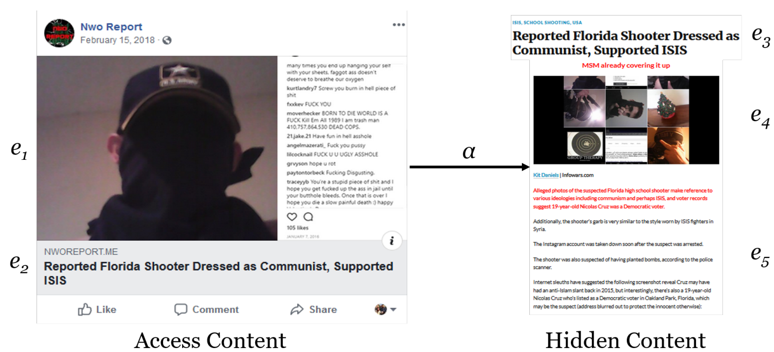

Fig. 1. Example of Access and Hidden Content

Figure 1 shows examples of the above definitions. It shows a Facebook post and the corresponding news article [3] that the post leads to. In this example, the post is an example of an Access Content that has two elements– e_1 (an image) and e_2 (the headline); in other words, $\mathcal{A} = \{e_1, e_2\}$. Once a user clicks on any $e \in \mathcal{A}$, it leads to the hidden content that has three elements- the headline (e_3), the image (e_4), and the text (e_5). So, $\mathcal{H} = \{e_3, e_4, e_5\}$.

An element may contain claims and/or evidences. A claim is a statement of opinion that the author is asking her or his audience to accept as true and an

[3] https://nworeport.me/2018/02/15/reported-florida-shooter-dressed-as-communist-supported-isis

evidence is the fact, data, or reasoning upon which the claim is based [50]. In general, elements in access contents contain claims; the expectation of readers is that the elements in hidden contents will contain evidences to support the claims. For instance, in Figure 1, e_2 claims that the shooter *'supported ISIS'*. Some claims can be explicit in an element but some claims may be implicit and can only be inferred from an element. For example, a user who has been aware of the **2018 Florida Parkland School Shooting** incident, may infer that the headline and the image is referring to the shooter *Nikolas Cruz*. Some claims can be obvious but still implicit. For example, the word *Florida* refers to a state of the United States. We define the set of all such claims, explicit or implicit, as the *Claim Context* of an element and denote with $\mathcal{C}_e$. Formally, if e holds n claims, we write, $\mathcal{C}_e = \{c_1, c_2, ..., c_n\}$, where c refers to one claim. Similarly, we define the set of evidences from an element as *Evidence Context*. If an element contains n evidences, we write, $\mathcal{V}_e = \{v_1, v_2, ..., v_n\}$, where v refers to one evidence. This definition allows the contexts (claim or evidence) to be built from knowledge networks, databases, and other data sources of similar nature through query systems as well as from the element itself through information extraction.

Definition 5. *Misleading Element Pair: Given two elements, $e_i \in \mathcal{A}$ and $e_j \in \mathcal{H}$, where $\mathcal{A}\overrightarrow{\alpha}\mathcal{H}$, if $\exists\ c \in \mathcal{C}_{e_i}$ and $\exists!\ v \in \mathcal{V}_{e_j}$ such that v substantiates c,* ***or*** *$\exists\ c \in \mathcal{C}_{e_i}$ and $\exists\ v \in \mathcal{V}_{e_j}$ such that c misrepresents v, we say that e_i and e_j forms a misleading element pair and denote as $e_i \rightsquigarrow e_j$.*

For example, in Figure 1, e_2 claims that the shooter supported ISIS but e_5 doesn't have any evidence that substantiates the claim. So, according to the above definition, $e_2 \rightsquigarrow e_5$. Later, we describe different instances of this definition with appropriate real-world examples.

Definition 6. *Misleading Connection: Given $\mathcal{A}$ and $\mathcal{H}$, where $\mathcal{A}\overrightarrow{\alpha}\mathcal{H}$, if $\exists\ \mathcal{E}$ such that $\mathcal{E} \neq \emptyset$ and $\mathcal{E} = \{(e_i, e_j) : e_i \in \mathcal{A}$ and $e_j \in \mathcal{H}$ and $e_i \rightsquigarrow e_j\}$, we say that $\mathcal{A}$ is a misleading connection to $\mathcal{H}$ with respect to $\mathcal{E}$ and denote as $\mathcal{A} \rightsquigarrow_{\mathcal{E}} \mathcal{H}$.*

For instance, in Figure 1, we say $\mathcal{A} \rightsquigarrow_{\{(e_2,e_5)\}} \mathcal{H}$, assuming $e_2 \rightsquigarrow e_5$ is the only misleading element pair.

2.3 Different Types of Misleading Connection

In this section, we use the above definitions to categorize misleading connection, formally characterize each of them, and explain with real-world examples.

Unsubstantiated Claim: It is a category of misleading connection where the access point contains one or more claims that are not substantiated by evidence in the hidden content. Formally, a misleading connection $\mathcal{A} \rightsquigarrow_{\mathcal{E}} \mathcal{H}$ belongs to this category if $\exists\ e_i \in \mathcal{A}$ such that $\forall e_j \in \mathcal{H},\ (e_i, e_j) \in \mathcal{E}$. For instance, in Figure 1, e_2 is not substantiated by any $e \in \mathcal{H}$. So, $\{(e_2, e_3), (e_2, e_4), (e_2, e_5)\} \subseteq \mathcal{E}$ where $\mathcal{A} \rightsquigarrow_{\mathcal{E}} \mathcal{H}$. It is an example of an unsubstantiated claim. Note that, even though e_2 and e_3 are same, $\mathcal{V}_{e_3}$ doesn't contain any evidence that supports the claim 'supported ISIS'. So, according to Definition 5, $e_2 \rightsquigarrow e_3$.

Exaggeration: A misleading connection $\mathcal{A} \rightsquigarrow_{\mathcal{E}} \mathcal{H}$ is an exaggeration if $\exists\ (e_i, e_j) \in \mathcal{E}$ such that a claim $c \in \mathcal{C}_{e_i}$ exaggerates an evidence $v \in \mathcal{V}_{e_j}$. For instance, a New York Times article [4] has the headline *China's Censors Ban Winnie the Pooh and the Letter 'N' After Xi's Power Grab*. It claims that the letter 'N' has been 'banned' by Chinese Censors after President Xi's power grab. However, the only pertaining evidence in the article body says that the letter 'N' was just 'censored' for a while. So, we argue that it is an example of exaggeration. Generally, we see the use of clickbait [39] to exaggerate by exploiting sensational, vulgar, or attention seeking words but it is also possible to exaggerate without using a clickbait as this example shows.

Hiding Context: In this type of misleading connection, access point misses one or more important claims and this absence makes the access point misrepresenting with respect to the hidden context. By formal definition, hiding context is a misleading connection $\mathcal{A} \rightsquigarrow_{\mathcal{E}} \mathcal{H}$, where $\exists e_j \in \mathcal{H}$ such that $\exists v \in \mathcal{V}_{e_j}$, which associated claim $c \notin e_i \in \mathcal{A}$. This type of omission only reveals partial information which may create biasness among the readers. Example 2 shows this type of misleading connection that hides the context for creating biasness among the readers. The headline puts its focus on 21,000 automated tweets that consistently supported Labour party during last election. But from the context, readers can also found that 13,000 automated tweets were also discovered supporting the Conservative Party. As people of modern age like to consume most of the news from headlines only, this type of headline may create a wrong impression on users' mind.

Example 2. Labour election campaign boosted by fake Twitter accounts [5]

Impersonation: Impersonation type misleading connection happens when the implicit claims of the access points are misused to attract the readers. More precisely, the content creators impersonate any famous/popular entity (place, person, organization, etc.) on the headline/image to report an incident, but the hidden content represents other entity. Formally, a misleading connection $\mathcal{A} \rightsquigarrow_{\mathcal{E}} \mathcal{H}$ is impersonating type if $\exists\ (e_i, e_j) \in \mathcal{E}$ such that e_i is an implicit claim and the $\exists v \in \mathcal{V}_{e_j}$ which alternates or changes the perception of the claim, e_i. For anyone scrolling through the social media, the headline showed in Example 3 is likely to have been taken as talking about the Brexit campaigner, Gina Miller [6], unless they click the full story. Sometimes reporters use fast forward referencing techniques with visuals of the actor who is not the protagonist to trick the readers which is another type of misleading connection technique.

Example 3. Gina Miller jailed after conning clients out of 800,000 to fund lavish life[7]

[4] https://www.nytimes.com/2018/02/28/world/asia/china-censorship-xi-jinping.html
[5] https://www.telegraph.co.uk/news/2017/06/01/exclusive-labour-election-campaign-boosted-fake-twitter-accounts/
[6] https://en.wikipedia.org/wiki/Gina_Miller
[7] https://tinyurl.com/y37lycz3

There are many other instances of misleading connection are available. For example, important contextual information can be omitted to mislead the readers as we showed in the section 1. Sometimes, omitting such information can create the biasness among users. Moreover, impersonating (where the identity of a famous person misused), the claim made in the access element contradicted by the evidence presented in the hidden content [8], use of equivocating access elements are some other types of misleading connections.

2.4 Misleading Connection in Social Media

Although misleading connection problem spans over different online platforms (e.g. online news sites, online review sites, search engine result page, web pages, etc.), we will analyze the nature of misleading connection on social media only where the problem is more severe and more diverse. Analyzing the social media platforms will be useful in understanding the different dynamics of the problem and discovering discriminating features, some of which we believe can also be applicable to other platforms.

As a source of information, social media platforms are getting much higher attention in recent times [43]. Media organizations also find this platform useful to disseminate their news to a larger audience faster [54]. There are many ways a misleading connection can be established in social media. As news media target social media as a primary tool for disseminating news, it is not surprising that some media organizations deliberately create misleading connection between the access content and hidden content of the news articles to attract more readers. News producers try to construct the headline as much as catchy to grab the attention of the readers and in doing so sometimes they omit critical information or exaggerate the event. As a result, the impression created from the headline can go wrong and the wrong impression can even be lasted after reading the full article. In the example showed in figure 1, the deliberate inclusion of the claim "supported ISIS" without the proper evidence presenting in the content is the representative of the described misleading connection. Moreover, to attract a particular group of readers news providers often customize their headlines depending on the users' race, location and news consumption patterns. Headlines showed in social media posts often include different version than the original article has [8]. This type of customization can often misrepresent the original content and thus mislead the readers.

Not only the headlines, the visuals used with the social media posts and news articles can also be used in establishing misleading collection. In many video streaming based social media sites, for example, YouTube, Vimeo, Daily Motion etc., content publishers deliberately employ techniques that aim to deceive viewers into clicking their videos by using of eye-catching thumbnails, such as depictions of abnormal stuff or attractive adults, which are often irrelevant to video content [60]. Figure 2 shows a YouTube video where the thumbnail is

[8] https://www.gurufocus.com/news/223102/when-headlines-contradict-content-

manipulated to trick the viewers and the thumbnail here creates a misleading connection.

There are some special characteristics of the social media which make the problem more critical. As users prefer to follow and form networks with like-minded people, the traditional methods of content distribution get affected by the peer-to-peer distribution of content which results in an echo chamber effect [20]. The echo chamber effect prevents users from exposing to alternative views and verified information. So, when social media posts containing misleading connections get viral, echo chamber effect can magnify the risk of misleading mass people.

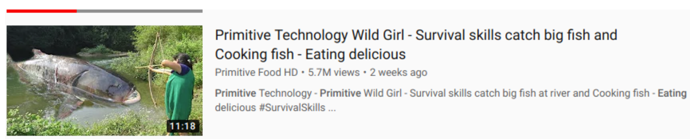

Fig. 2. Manipulated Video Thumbnail

3 *Misleading Connection* Detection

In this section, we will define misleading connection detection task formally based on the notations and definitions we introduced in Section 2.1.

Definition 7. *Misleading Connection Detection: The task of misleading connection detection is to identify if there exists an elements pair, (e_i, e_j) where $e_i \in \mathcal{A}$ and $e_j \in \mathcal{H}$, which is misleading given the claim context, $\mathcal{C}_{ei}$ and evidence context, $\mathcal{V}_{ej}$*

Mathematically,

$$\chi = \begin{cases} 1, & \textit{if } \exists\ c \in \mathcal{C}_{e_i} \textit{ and } \exists\ v \in \mathcal{V}_{e_j} \textit{ such that } c \textit{ misrepresents } v \\ 0, & \textit{otherwise} \end{cases} \quad (1)$$

where χ is the prediction function we need to learn.

4 Challenges

To combat misleading connection successfully, we need to identify the challenges associated with it. The data collection strategies, effective model selection, representations of features, etc. are the primary ingredients of this challenging task. Moreover, the unique characteristics of different domains and applications have added extra dimensions to the problem. In this section, we summarize them from different aspects of the problem.

4.1 Dataset Construction

Currently, there is no existing benchmark dataset that can serve our purpose fully. Yoon et al. [58] crawled a large corpus of news articles published in South Korea and automatically generated the labels by implanting the topically-inconsistent content into the body. But, we argue that the model built on dataset may identify the incongruency between the headline and the news body, but in real life scenario, the incongruency may not be created by different topical content only. Topically-consistent content can also create incongruency. Moreover, all techniques involved in establishing misleading connection, for example, misled by visual elements, are not addressed in this dataset. Another dataset curated by Wei et al. [55] covers the identification of ambiguous and misleading news headlines problem, but like the previous dataset, it doesn't include the other elements associated with misleading connection.

Zhang et al. [61] gathered a dataset of 40 highly shared articles focused on two topics: public health and climate science. This small dataset is annotated with c16 redibility indicators. Although the indicators are determined from analyzing the title and body of the articles only, other metadata which can also be used in misleading connection are ignored here. But these indicators can be a good starting point for exploring our problem from the textual context only.

As the misleading connection problem involves different types of elements like textual, visuals, etc. and spans over different domains, constructing a comprehensive dataset to cover all the aspects might be challenging. Obtaining a reliable dataset will be a very time and labor intensive because the process will require expert annotators to perform a careful analysis of the connecting contexts, domain-specific knowledge, and additional evidence.

4.2 Feature Representation

Different linguistic-based features have been widely studied for general NLP tasks, such as text classification [33, 16, 41], deception detection tasks [15] and also visual features have been used in manipulated image identification tasks [3], but the unique aspects of misleading connection problem may demand for more effective feature representations. For example, only a particular type of feature representation or mere combination of individual feature representations can't capture the underlying characteristics of the misleading connection when multiple access elements get involved. Moreover, automatic systems for misleading connection detection based on known features can potentially be fooled, and carefully crafted misleading connection may go undetected. Additionally, the dynamics of the online platform are changing over time which can redefine the dimensions of the problem. So, building an adaptive type of feature representations that can cope with the continuous changes can be a difficult task.

4.3 Model Construction

Misleading connection has different characteristics which make it different from other types of information disorder. So, to combat this problem effectively using data mining solutions, we have to address the following challenges which can be emerged from its unique characteristics.

As there is no single model that is better than all other models for all problems [56], the selection of a model is always been a challenging task. Some models may work well in one type of problem but may struggle in other tasks. For example, recursive neural network (RNN) founds to be effective for sequence learning [47, 35], where convolutional neural network (CNN) shows efficacy in spatial pattern learning [25]. So, choosing an effective model covering all the aspects of the problem might be a challenging task.

Data representation is a crucial part of designing data mining model and has an important impact on the performance of the model [44]. The relationships between the elements of the access content and hidden content can be conceptualized through the different representation techniques. For example, we can pair the headline and the whole content of the news article or we can split the content by sentences and make a set of headline-sentence pair to feed into the model.

Input data can be both static and streaming in nature. The model built on static data, for example, news headline and content, will be inefficient for handling streaming data like video. Moreover, the data processing speed of the static data and streaming data can differ largely. The model which can decide the quality of the connection in the run time may face much delay while applied on the streaming type data.

Interpretability of the model will help the users understand the reasons behind the decision it takes and make the decision trustworthy. For example, if a non-interpretable model built on word features and visual features, declares that a news article has a misleading connection, users might not be able to understand which components or attributes are mainly responsible for the decision. Although deep learning models are proven to be very much effective for the detection tasks, they are often used as a black box and lack of interpretability [48]. So building an effective neural based model with interpretability will be a challenging task.

5 Possible Solutions and Research Directions

5.1 Dataset Related

Combating misleading connection requires the involvement of every factor associated with it. So, a rich and comprehensive dataset containing the attribution of different components like contextual information of the elements (e.g. headline, image, content), platform related metadata (e.g. posts, reactions, sharing pattern), etc. is required. A promising direction is to create a comprehensive and large-scale misleading connection benchmark dataset, which can be used by researchers to facilitate further research in this area.

5.2 Feature Extraction Related

In this subsection, we provide some outlines to buildup different feature representations based on the type of the elements of access content, hidden content, and domain related information.

Misleading connections that involve textual type access point elements use deceptive techniques to mislead people. Thus it is reasonable to explore different types of linguistic features that can be extracted from the content. In natural language processing, common linguistic features are often used to represent a document for classification task [33, 16, 41] and can be categorized into - i) Lexical Features (related to structure of the word e.g. Total words, character per words, etc. [31]) , ii) Syntactic Features (represents structure of the sentences e.g. The frequency of words and phrases, n-grams, bag-of-words and TFIDF representation, POS Tagging etc. [36, 18]), iii) Semantic Features (explores conceptual patterns of the document e.g. use of synonym, antonym, multi meaning expressive words etc. [4]). In addition to that, stylistic feature representation to capture the cues related to writing styles such as vocabulary complexity(number of big words), grammatical complexity (use of demonstrative adjective, conjunction, etc), uncertainty (number of modal verbs), forward-referencing, etc . can be useful [1, 5].

Visual features extracted from images are shown to be important indicators for other information disorder like fake news [28]. Moreover, Images in social media posts can also be utilized to better understand users' sentiments [51] toward news events. So, it's imperative to explore the visual features that can contribute to misleading connection identification tasks. There are some existing works which use image metadata to verify the image integrity in online [13, 34, 22] but none of them exploit effective visual features [49] which can more effective in information disorder detection task. So, exploring the effectiveness of the feature representations extracted from images can be a potential research direction.

Domain related features can be captured from the particular information related to domain's activity, identity etc. For example, in news media, the credibility of the news article is associated with the number of quotations and hyperlinks it uses [9, 46]. Dhoju et al. [10] used the number of hyperlinks, social media mention, and other linguistic feature to predict the source (reliable or unreliable) of health related news articles. Although misleading connection can be found on different platforms (e.g. online news sites, social media, online review sites, etc.) for space limitation, we will only focus on social media platform. Social media is the most popular platform for the users to news consumption and dissemination. So, social media related features can also be derived from the user-driven social engagements of information consumption. For example, users' profile metadata [7], social media groups related data [57, 29, 30], status/post related information [40, 27, 30, 7] can be considered for our tasks. Moreover, network based feature set [23, 26] can be effective in understanding the propagation techniques of contents having misleading connections.

5.3 Identification Model Related

We will discuss some possible method constructions processes based on the features and characteristics we analyzed before. Specifically, we will propose some models exploring various aspects of the problem, and analyzing existing work in other areas. Note that, the suggested methodologies may not address all the misleading connection types but can solve some of them.

One way to identify where the headline is misleading with respect to the content could be generating a statistically best headline for an article using existing title and headline generation and summarisation methods[2, 11, 59] and evaluate how far away the existing headline is from this in terms of a number of criteria, such as lexical choices, syntactic structure, length, tonality (sentiment or emotion), and so on.

One type of misleading connection can be established when an element of the access content holds a claim but the evidence of the claim is absent in the hidden content. So argument mining task can be influential in identifying this type of misleading connection. For example, Stab et al. [45] explore methods for the identification of arguments supported by insufficient evidence.

In addition to the analysis of content, other works and systems, focus on the use of network analysis or propagation techniques to detect various information disorders [38, 42, 21, 17]. The studies show that different diffusion patterns exist that characterise misinformation vs. legitimate memes, with misinformation patterns propagating in a more viral way [17] and often being generated by bots and not humans [38]. These type of models can also be constructed for identifying and monitoring misleading connection.

In addition to the above-mentioned methodologies, there are some other techniques which can be incorporated in misleading connection detection task. Stance detection [32] of the headline-news content can be a good technique to identify the incongruency between them. Moreover, the stance of the comments put by the users can also be an indicative factor of the problem. Various deception detection techniques in writing [14] can be useful in identifying deceptive access content. Models building on the temporal patterns difference between credible and misleading content [27] can be also useful for misleading connection detection.

6 Preliminary Experiment

To understand the complexity and characteristics of misleading connection problem, we conduct a preliminary study with randomly picked 50 articles (25 from mainstream media and 25 from unreliable media, and this media categorization is described in [39]). We splitted each articles by the paragraph and made headline-paragraph pairs and worked with such 315 pairs and checked whether the headline is misleading with respect to the content. We used 3 labels : Misleading, Not Misleading, Not Sure. If the headline is found to be misleading, we identified the misleading type. The available misleading types were : i) Exaggerates, ii) Equivocates, iii) Hides Context, iv) Adds Extra Context, v) Contradicts,

vi) Impersonates, vii) Cherry Picks , and iv) Something Else. Our preliminary study found only 10 misleading pairs of headline-paragraph and among them 4 of them were exaggerating, 3 of them were adding extra content, 2 were hiding context and only one was contradicts type of misleading. This study proved our hypothesis that the misleading connection problem does exist. Moreover, the annotation process also revealed the complexity of data annotation which requires good understanding of the problem and considerable amount of efforts.

7 Discussion

In this work, we have investigated the overview of the misleading connection problem, formally define it and analyze its characteristics that can be helpful in battling the problem. Moreover, we also suggest some outlines for data collection, feature extraction and designing effective solutions from data mining perspectives. Misleading connection detection is an unexplored research area. Our study reveals the challenges and potential research directions which focus on building effective and comprehensive solutions. As various circumstances like the human, societal and technological associated with the problem intensify the complexity of the problem, there is a demand of multidisciplinary research approach to design and develop methodologies, and technologies to effectively combat this malpractice. The solution to this problem has a broader impact in the long run. The problem is not confined in some particular domains only. Different online sites including product review, search engine result page, electronic communication like Email (e.g. Subject of the email can be misleading) are also affected by the problem. And the successful tackling of the current problem focusing on the social media platform can open the solution paths for other domains also.

8 Conclusion

With the advancement of technologies and various media platforms, the problem with the quality of information is also increasing. In this paper, we introduce a new type of information disorder which has a clear negative impact on the netizens. We also give a brief overview of the problem, identify its characteristics, suggested possible solution techniques and discuss challenges with potential research scopes from the data mining perspectives. Moreover, we also analyze a number of existing datasets, but none of them are appropriate for our task, which sets up the groundwork for the next steps of our future work.

References

1. Afroz, S., Brennan, M., Greenstadt, R.: Detecting hoaxes, frauds, and deception in writing style online. In: 2012 IEEE Symposium on Security and Privacy. pp. 461–475. IEEE (2012)

2. Banko, M., Mittal, V.O., Witbrock, M.J.: Headline generation based on statistical translation. In: Proceedings of the 38th Annual Meeting on Association for Computational Linguistics. pp. 318–325. Association for Computational Linguistics (2000)
3. Bayram, S., Avcibas, I., Sankur, B., Memon, N.D.: Image manipulation detection. Journal of Electronic Imaging **15**(4), 041102 (2006)
4. Bloehdorn, S., Hotho, A.: Boosting for text classification with semantic features. In: International workshop on knowledge discovery on the web. pp. 149–166. Springer (2004)
5. Blom, J.N., Hansen, K.R.: Click bait: Forward-reference as lure in online news headlines. Journal of Pragmatics **76**, 87–100 (2015)
6. Carvalho, C., Klagge, N., Moench, E.: The persistent effects of a false news shock. Journal of Empirical Finance **18**(4), 597–615 (2011)
7. Castillo, C., Mendoza, M., Poblete, B.: Information credibility on twitter. In: Proceedings of the 20th international conference on World wide web. pp. 675–684. ACM (2011)
8. Chesney, S., Liakata, M., Poesio, M., Purver, M.: Incongruent headlines: Yet another way to mislead your readers. In: Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism. pp. 56–61 (2017)
9. De Maeyer, J.: The journalistic hyperlink: Prescriptive discourses about linking in online news. Journalism Practice **6**(5-6), 692–701 (2012)
10. Dhoju, S., Rony, M.M.U., Hassan, N.: Differences between health related news articles from reliable and unreliable media. arXiv preprint arXiv:1811.01852 (2018)
11. Dorr, B., Zajic, D., Schwartz, R.: Hedge trimmer: A parse-and-trim approach to headline generation. In: Proceedings of the HLT-NAACL 03 on Text summarization workshop-Volume 5. pp. 1–8. Association for Computational Linguistics (2003)
12. Ecker, U.K., Lewandowsky, S., Chang, E.P., Pillai, R.: The effects of subtle misinformation in news headlines. Journal of experimental psychology: applied **20**(4), 323 (2014)
13. Elkasrawi, S., Dengel, A., Abdelsamad, A., Bukhari, S.S.: What you see is what you get? automatic image verification for online news content. In: 2016 12th IAPR Workshop on Document Analysis Systems (DAS). pp. 114–119. IEEE (2016)
14. Feng, S., Banerjee, R., Choi, Y.: Syntactic stylometry for deception detection. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2. pp. 171–175. Association for Computational Linguistics (2012)
15. Feng, V.W., Hirst, G.: Detecting deceptive opinions with profile compatibility. In: Proceedings of the Sixth International Joint Conference on Natural Language Processing. pp. 338–346 (2013)
16. Forman, G.: An extensive empirical study of feature selection metrics for text classification. Journal of machine learning research **3**(Mar), 1289–1305 (2003)
17. Friggeri, A., Adamic, L., Eckles, D., Cheng, J.: Rumor cascades. In: Eighth International AAAI Conference on Weblogs and Social Media (2014)
18. Fürnkranz, J.: A study using n-gram features for text categorization. Austrian Research Institute for Artifical Intelligence **3**(1998), 1–10 (1998)
19. Gabielkov, M., Ramachandran, A., Chaintreau, A., Legout, A.: Social clicks: What and who gets read on twitter? ACM SIGMETRICS Performance Evaluation Review **44**(1), 179–192 (2016)
20. Goldie, D., Linick, M., Jabbar, H., Lubienski, C.: Using bibliometric and social media analyses to explore the echo chamber hypothesis. Educational Policy **28**(2), 281–305 (2014)

21. Gupta, A., Kumaraguru, P., Castillo, C., Meier, P., et al.: Tweetcred: A real-time web-based system for assessing credibility of content on twitter. arXiv preprint arXiv:1405.5490 (2014)
22. Gupta, A., Lamba, H., Kumaraguru, P., Joshi, A.: Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy. In: Proceedings of the 22nd international conference on World Wide Web. pp. 729–736. ACM (2013)
23. Gupta, M., Zhao, P., Han, J.: Evaluating event credibility on twitter. In: Proceedings of the 2012 SIAM International Conference on Data Mining. pp. 153–164. SIAM (2012)
24. Ireton, C., Posetti, J.: Journalism,fake news& disinformation (2018)
25. Jaderberg, M., Simonyan, K., Zisserman, A., et al.: Spatial transformer networks. In: Advances in neural information processing systems. pp. 2017–2025 (2015)
26. Jin, Z., Cao, J., Jiang, Y.G., Zhang, Y.: News credibility evaluation on microblog with a hierarchical propagation model. In: 2014 IEEE International Conference on Data Mining. pp. 230–239. IEEE (2014)
27. Jin, Z., Cao, J., Zhang, Y., Luo, J.: News verification by exploiting conflicting social viewpoints in microblogs. In: Thirtieth AAAI Conference on Artificial Intelligence (2016)
28. Jin, Z., Cao, J., Zhang, Y., Zhou, J., Tian, Q.: Novel visual and statistical image features for microblogs news verification. IEEE transactions on multimedia **19**(3), 598–608 (2017)
29. Kwon, S., Cha, M., Jung, K., Chen, W., Wang, Y.: Prominent features of rumor propagation in online social media. In: 2013 IEEE 13th International Conference on Data Mining. pp. 1103–1108. IEEE (2013)
30. Ma, J., Gao, W., Wei, Z., Lu, Y., Wong, K.F.: Detect rumors using time series of social context information on microblogging websites. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. pp. 1751–1754. ACM (2015)
31. Melville, P., Gryc, W., Lawrence, R.D.: Sentiment analysis of blogs by combining lexical knowledge with text classification. In: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 1275–1284. ACM (2009)
32. Mohammad, S., Kiritchenko, S., Sobhani, P., Zhu, X., Cherry, C.: Semeval-2016 task 6: Detecting stance in tweets. In: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). pp. 31–41 (2016)
33. Moschitti, A., Basili, R.: Complex linguistic features for text classification: A comprehensive study. In: European Conference on Information Retrieval. pp. 181–196. Springer (2004)
34. Pasquini, C., Brunetta, C., Vinci, A.F., Conotter, V., Boato, G.: Towards the verification of image integrity in online news. In: 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). pp. 1–6. IEEE (2015)
35. Peters, M.E., Ammar, W., Bhagavatula, C., Power, R.: Semi-supervised sequence tagging with bidirectional language models. arXiv preprint arXiv:1705.00108 (2017)
36. Post, M., Bergsma, S.: Explicit and implicit syntactic features for text classification. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). vol. 2, pp. 866–872 (2013)
37. Potthast, M., Köpsel, S., Stein, B., Hagen, M.: Clickbait detection. In: European Conference on Information Retrieval. pp. 810–817. Springer (2016)

38. Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Patil, S., Flammini, A., Menczer, F.: Truthy: mapping the spread of astroturf in microblog streams. In: Proceedings of the 20th international conference companion on World wide web. pp. 249–252. ACM (2011)
39. Rony, M.M.U., Hassan, N., Yousuf, M.: Diving deep into clickbaits: Who use them to what extents in which topics with what effects? In: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017. pp. 232–239. ACM (2017)
40. Ruchansky, N., Seo, S., Liu, Y.: Csi: A hybrid deep model for fake news detection. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. pp. 797–806. ACM (2017)
41. Scott, S.: Feature engineering for a symbolic approach to text classification. University of Ottawa (Canada) (1998)
42. Seo, E., Mohapatra, P., Abdelzaher, T.: Identifying rumors and their sources in social networks. In: Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR III. vol. 8389, p. 83891I. International Society for Optics and Photonics (2012)
43. SHEARER, E.: Social media outpaces print newspapers in the u.s. as a news source. https://www.pewresearch.org/fact-tank/2018/12/10/social-media-outpaces-print-newspapers-in-the-u-s-as-a-news-source/ (2018)
44. Siraj, F., Omer, E.A.O.A., Hasan, M.R.: Data mining and neural networks: the impact of data representation. In: Advances in Data Mining Knowledge Discovery and Applications. IntechOpen (2012)
45. Stab, C., Gurevych, I.: Recognizing insufficiently supported arguments in argumentative essays. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers. pp. 980–990 (2017)
46. Sundar, S.S.: Effect of source attribution on perception of online news stories. Journalism & Mass Communication Quarterly **75**(1), 55–68 (1998)
47. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. In: Advances in neural information processing systems. pp. 3104–3112 (2014)
48. Tan, S., Sim, K.C., Gales, M.: Improving the interpretability of deep neural networks with stimulated learning. In: 2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU). pp. 617–623. IEEE (2015)
49. ping Tian, D., et al.: A review on image feature extraction and representation techniques. International Journal of Multimedia and Ubiquitous Engineering **8**(4), 385–396 (2013)
50. Toulmin, S.E.: The uses of argument. Cambridge university press (2003)
51. Wang, Y., Wang, S., Tang, J., Liu, H., Li, B.: Unsupervised sentiment analysis for social media images. In: Twenty-Fourth International Joint Conference on Artificial Intelligence (2015)
52. Wardle, C.: Fake news. its complicated. https://firstdraftnews.org/fake-news-complicated/ (2017)
53. Wardle, C., Derakhshan, H.: Information disorder: Toward an interdisciplinary framework for research and policy making. Council of Europe report, DGI (2017) **9** (2017)
54. Weeks, B.E., Holbert, R.L.: Predicting dissemination of news content in social media: A focus on reception, friending, and partisanship. Journalism & Mass Communication Quarterly **90**(2), 212–232 (2013)

55. Wei, W., Wan, X.: Learning to identify ambiguous and misleading news headlines. arXiv preprint arXiv:1705.06031 (2017)
56. Wolpert, D.H.: The lack of a priori distinctions between learning algorithms. Neural computation **8**(7), 1341–1390 (1996)
57. Yang, F., Liu, Y., Yu, X., Yang, M.: Automatic detection of rumor on sina weibo. In: Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics. p. 13. ACM (2012)
58. Yoon, S., Park, K., Shin, J., Lim, H., Won, S., Cha, M., Jung, K.: Detecting incongruity between news headline and body text via a deep hierarchical encoder. arXiv preprint arXiv:1811.07066 (2018)
59. Zajic, D., Dorr, B., Schwartz, R.: Automatic headline generation for newspaper stories. In: Workshop on Automatic Summarization. pp. 78–85. Citeseer (2002)
60. Zannettou, S., Chatzis, S., Papadamou, K., Sirivianos, M.: The good, the bad and the bait: Detecting and characterizing clickbait on youtube. In: 2018 IEEE Security and Privacy Workshops (SPW). pp. 63–69. IEEE (2018)
61. Zhang, A.X., Ranganathan, A., Metz, S.E., Appling, S., Sehat, C.M., Gilmore, N., Adams, N.B., Vincent, E., Lee, J., Robbins, M., et al.: A structured response to misinformation: Defining and annotating credibility indicators in news articles. In: Companion of the The Web Conference 2018 on The Web Conference 2018. pp. 603–612. International World Wide Web Conferences Steering Committee (2018)