

LITERATURE SURVEY

1. TITLE : Semi- Automatic Classification and Duplicate Detection From Human Loss News Corpus

AUTHOR : Adnan Abid, Waqas Ali, Muhammad Shoaib Farooq

YEAR : 2020

Automatic news repository collection systems involve a news crawler that extracts news from different news portals, subsequently, these news need to be processed to figure out the category of a news. article e.g. sports, politics, showbiz etc. In this process there are two main challenges first one is to place a news article under the right category of news, while the second one is to detect a duplicate news, i.e. when the news are being extracted from multiple sources, it is highly probable to get the same news from many different portals, resulting into duplicate news; failing to which may result into inconsistent statistics obtained after pre-processing the news text. This problem becomes more pertinent when we deal with human loss news involving crime, accident etc. related news articles. As the system may count the same news many times resulting into misleading statistics. In order to address these problems, this research presents the following contributions. Firstly, a news corpus comprising of human loss news of different categories has been developed by gathering data from different sources of well-known and authentic news websites. The corpus also includes a number of duplicate news. Secondly, a comparison of different classification approaches has been conducted to empirically find out the best suitable text classifier for the categorization of different sub-categories of human loss news.

Lastly, methods have been proposed and compared to detect duplicate news from the corpus by involving different pre-processing techniques and widely used similarity measures, cosine similarity, and Jaccard's coefficient. The results show that conventional text classifiers are still relevant and perform well in text classification tasks as MNB has given 89.5% accurate results. While, Jaccard coefficient exhibits much better results than Cosine similarity for duplicate news detection with different pre-processing variations with an average accuracy of 83.16%.

2. TITLE : News Information Platform Optimization Based on the Internet of Things

AUTHOR : Hongyun Tan, Yiping Li

YEAR : 2021

The Internet of Things device online recommendation system has been applied in some Internet of Things operating companies and has achieved good results. In the process of designing and implementing the Internet of Things equipment online promotion system, this article uses the news information protocol transmission structure language to explain the use case analysis and activity diagram analysis of the Internet of Things equipment online promotion system and uses the Spring Hibernate (SH) integration framework in the field of the news information topology layer under the Internet of Things.

Then, design and implement the technical architecture and main functional modules of the Internet of Things device online recommendation system, effectively improving the development efficiency and operating quality of the Internet of Things device online recommendation system. The system design concept and implementation ideas can be used as a reference for related industries when developing enterprise applications. This article uses the news information protocol transmission structure language to explain the demand analysis of the Internet of Things equipment online promotion system and mainly discusses the analysis of the use case of the Internet of Things equipment online promotion system and the analysis of core business activity diagrams. We completed the design of the Internet of Things device online promotion system based on the news information topology layer platform under the Internet of Things, mainly using the integration framework in the field of the news information topology layer under the Internet of Things to design the technical architecture of the Internet of Things device online promotion system, and design the Internet of Things of the function module structure and data table structure of the equipment online recommendation system. At the same time, they complete the realization of the main functions of the Internet of Things device online recommendation system; elaborate on the realization process of core modules such as Internet of Things device category management, Internet of Things device information management, announcement management, and device recommender management; and discuss the system testing process and application effects.

3. TITLE : Fake News Outbreak 2021: Can We Stop The Viral spread?

AUTHOR : Tanveer Khan, Antonis Michala's, Adnan Akhundzada

YEAR : 2021

Social Networks' omnipresence and ease of use has revolutionized the generation and distribution of information in today's world. However, easy access to information does not equal an increased level of public knowledge. Unlike traditional media channels, social networks also facilitate faster and wider spread of disinformation and misinformation. Viral spread of false information has serious implications on the behavior's, attitudes and beliefs of the public, and ultimately can seriously endanger the democratic processes. Limiting false information's negative impact through early detection and control of extensive spread presents the main challenge facing researchers today. In this survey paper, we extensively analyze a wide range of different solutions for the early detection of fake news in the existing literature. More precisely, we examine Machine Learning (ML) models for the identification and classification of fake news, online fake news detection competitions, statistical outputs as well as the advantages and disadvantages of some of the available data sets. Finally, we evaluate the online web browsing tools available for detecting and mitigating fake news and present some open research challenges.

4. TITLE : Data Acquisition Method of Sensor News Based on Collaborative Filtering Algorithm

AUTHOR : Jin Tong, Jian Sun

YEAR : 2022

With the vigorous development of new media technologies such as Internet of things, big data, and cloud computing, data-based sensor news (SN) will become the trend of news reporting in the future and the new normal of news production. Under this background, this paper further analyzes the relationship between SN production mode and traditional news production, including the inheritance of traditional news production value concept, as well as the breakthrough and change in form, media, and effect. In this paper, collaborative filtering (CF) algorithm is improved to solve the problems of data sparseness, user interest migration, and scalability in CF technology. In the calculation of news content similar degree (SD), the influence of part of speech and position of feature words in news is also considered, and the time window is used to establish a model that adapts to the change of user interest with time. In this method, the contribution degree of different attributes to distinguishing users is considered, and the attribute SD between users is accurately calculated, which effectively improves the accuracy of SN data acquisition results.

5. TITLE : Application Research of Fake News and Rumors Detection in Complex Network Environment

AUTHOR : Meng – Zhe Huang, Rong –Wang Yin

YEAR : 2022

Nowadays, the network environment is very complicated, and so is the information transmission in the network. False news and rumors have become a big problem in the network environment. How to detect the elective information content in the complex network environment? Interest in elective detection techniques has also grown rapidly in recent years. -ere is an urgent need to develop elective tools to address this challenge by employing advanced Artificial Intelligence (AI) technologies. In this article, we analyze and study the current state of fake news and rumors in the complex network environment, summarize different methods of detecting fake news and rumors, and point out the important directions for the application of intelligent models in the detection of false information sources. -e main purpose is to show possible solutions on the one hand, and on the other hand to determine the main challenges and methodological inadequacy to stimulate future research.

6. TITLE : A Taxonomy of Fake News Classification Techniques: Survey and Implementation Aspects

AUTHOR : Dhiren, Keyur Patel, Urvish Thakker, Sudeep Tanwar, Rajesh

YEAR : 2022

In the present era, social media platforms such as Facebook, WhatsApp, Twitter, and Telegram are significant sources of information distribution, and people believe it without knowing their origin and genuineness. Social media has fascinated people worldwide in spreading fake news due to its easy availability, cost-effectiveness, and ease of information sharing. Fake news can be generated to mislead the community for personal or commercial gains. It can also be used for other personal benefits such as defaming eminent personalities, amendment of government policies, etc. Thus, to mitigate the awful consequences of fake news, several research types have been conducted for its detection with high accuracy to prevent its fatal outcome. Motivated by the aforementioned concerns, we present a comprehensive survey of the existing fake news identification techniques in this paper. Then, we select Machine Learning (ML) models such as Long Short Term Memory (LSTM), Passive Aggressive Algorithm, Random Forest (RF), and Naïve Bayes (NB) and train them to detect fake news articles on the self-aggregated dataset. Later, we implemented these models by hyper tuning various parameters such as smoothing, drop out factor, and batch size, which has shown promising results in accuracy and other evaluation metrics such as F1-score, recall, precision, and Area under the ROC Curve (AUC) score.

The model is trained on 6335 news articles, with LSTM showing the highest accuracy of 92.34% in predicting fake news and NB were showing the highest recall. Based on these results, we propose a hybrid fake news detection technique using NB and LSTM. At last, challenges and open issues along with future research directions are discussed to facilitate the research in this domain further.

7. TITLE : Are You a Cyborg, Bot or Human? - A Survey on Detecting Fake News Spreaders

AUTHOR : Wajiha Shahid, Yiran LI, Dakota Staples, Gulshan Amin, Saqib Hakak

YEAR : 2022

One of the major components of Societal Digitalization is Online social networks (OSNs). OSNs can expose people to different popular trends in various aspects of life and alter people's beliefs, behaviors, and decisions and communication. Social bots and malicious users are the significant sources for spreading misinformation on social media and can pose serious cyber threats in society. The degree of similarity of user profiles of a cyber bot and a malicious user spreading fake news is so great that it is very difficult to differentiate both based on their attributes. Over the years, researchers have attempted to find a way to mitigate this problem. However, the detection of fake news spreaders across OSNs remains a challenge. In this paper, we have provided a comprehensive survey of the state of art methods for detecting malicious users and bots based on different features proposed in

our novel taxonomy. We have also aimed to avert the crucial problem of fake news detection by discussing several key challenges and potential future research areas to help researchers who are new to this field.

8. TITLE : Effects of Font Style and Font Color in News Text on User Cognitive Load in Intelligent User Interfaces

AUTHOR : Xianglin Miao, Feijuan He, Yalin Miao

YEAR :2022

The advances on the Internet and news media technologies suggest intelligent and personalized media interfaces in order to improve reading efficiency of news readers and enhance news disseminations. Cognitive load is one of factors that affect the understanding ability of news readers and therefore news dissemination. It is expected that the display elements in an intelligent user interface of news media could be automatically adjusted to modulate the audience's perceived cognitive load level in order to improve the news reading efficiency. While the font style and font color of news text are the main display elements in the user interface of news media, it is not clear how these elements affect perceived cognitive load of news readers. This paper investigates perceived cognitive load of news readers under different font style and font color conditions. Experiments with the news text in Chinese as a case study found that the change of text font style of keywords resulted in the increase of the reader's perceived cognitive load during reading news text under the low introduced cognitive load and blue color displayed keywords.

While under the high introduced cognitive load, the italic font and red color of keywords text resulted in the decrease of the reader's perceived cognitive load significantly during reading news text, which therefore improves the news reading efficiency. This paper is limited to the effects of two aspects of news text (font styles and colors) on user's perceived cognitive load. The future work will focus on the investigation of effects of other factors such as news pictures and their colors, numbers, and locations on user's perceived cognitive load.

9. TITLE : A Graphical Decomposition and Similarity Measurement Approach for Topic Detection From Online News

AUTHOR : Kejing Xiao

YEAR : 2021

Topic detection aims to discover valuable topics from the massive online news. It can help people to capture what is happening in real world and alleviate the burden of information overload. It also has great significance since the online news is experiencing an explosive growth. Topic detection is typically transformed into a document clustering problem, whose core idea is to cluster news documents that report on the same topic to the same group based on document similarity. Due to the complex structure and long length of news documents, the similarity measurement of news is very challenging. Existing term-based methods represent news documents based on a set of informative keywords in the document with a vector space model (VSM) and then the relationship between documents is calculated by cosine similarity. However, VSM ignores the relationship between words and has sparse semantics, which leads to low precision of topic detection. In recent years, the probabilistic methods and the graph analytical methods have been proposed for topic detection.

However, both of them have high time complexity. To cope with these problems, we first present a novel document representation approach based on graphical decomposition, which decomposes each news document into different semantic units and then relationship between the semantic units is constructed to form a capsule semantic graph (CSG). The CSG can retain the relationship between words and alleviate the sparse semantics compared to VSM representation. We next introduce the graph kernel to measure the similarity between the CSGs based on their substructures. Finally, we use an incremental clustering method to cluster the news documents, in which the documents are represented by CSGs and the similarity between documents is calculated by graph kernel. The experiment results on three standard datasets show that our method obtains higher precision, recall and F1 score than several state-of-the-art methods. Moreover, the experiment results on a large news dataset show that our CSG-SM has lower time complexity than probabilistic methods and graph analytical methods.

10. TITLE : Effects of Incidental Brief Exposure to News on News Knowledge While Scrolling Through Videos

AUTHOR : Masanori Takano, Yuki Ogawa, Fumiaki Taka, Soichiro Morishita
YEAR : 2021

Increasing media choices due to online media diversification ensure that people without any interest in news avoid news media. This obstructs the construction of a shared social reality given the presence of politics news seekers and news avoiders. For mitigating these issues using media, incidental exposure to news on the Internet can be a powerful tool because it can bring news to the awareness of people who are politically disinterested.

We studied the effects of the glimpsing a news screen for less than a few seconds while watching online television news, termed incidental brief exposure, on news knowledge. For evaluating the effects, we combined the logs of news-watching behavior on an online television (for incidental brief exposure) and the results of a questionnaire survey (for news knowledge and media repertoires). We found that this incidental brief exposure mitigated the negative effect of social media usage on news knowledge. Although people with heavy social media usage have low news knowledge, heavy social media users with high frequently incidental brief exposure have more news knowledge than heavy social media users. As a possible scenario, memorizing news keywords due to incidental brief exposure may facilitate reading news related to these keywords when users incidentally encounter news on social media. On the other hand, the exposure did not moderate the effects of news media usage, such as mass media, curation sites, and online news sites. These findings suggest that incidental brief exposure while scrolling through videos, which is hardly noticed by users, enhances passive exposure effects in non-news media, such as social media