The Information Retrieval Series

Asif Ekbal
Rina Kumari

Dive into Misinformation Detection

From Unimodal to Multimodal and Multilingual Misinformation Detection

Springer

The Information Retrieval Series

Volume 30

Information Retrieval (IR) deals with access to and search in mostly unstructured information, in text, audio, and/or video, either from one large file or spread over separate and diverse sources, in static storage devices as well as on streaming data. It is part of both computer and information science, and uses techniques from e.g. mathematics, statistics, machine learning, database management, or computational linguistics. Information Retrieval is often at the core of networked applications, web-based data management, or large-scale data analysis.

The Information Retrieval Series presents monographs, edited collections, and advanced text books on topics of interest for researchers in academia and industry alike. Its focus is on the timely publication of state-of-the-art results at the forefront of research and on theoretical foundations necessary to develop a deeper understanding of methods and approaches.

This series is abstracted/indexed in EI Compendex and Scopus.

Asif Ekbal • Rina Kumari

Dive into Misinformation Detection

From Unimodal to Multimodal and Multilingual Misinformation Detection

Springer

Asif Ekbal
Computer Science and Engineering
Indian Institute of Technology Patna
Patna, India

Rina Kumari
KIIT University
Bhubaneswar, India

ISSN 1871-7500 ISSN 2730-6836 (electronic)
The Information Retrieval Series
ISBN 978-3-031-54833-8 ISBN 978-3-031-54834-5 (eBook)
https://doi.org/10.1007/978-3-031-54834-5

This book is dedicated to our parents who enlightened us with the value of strong work ethic and our families who always motivated for taking new challenges.

Preface

This book provides a brief introduction to misinformation and various novel approaches for solving misinformation detection problems. In recent times, most people get information through social media platforms. Social media news and information contain several attributes like text, image, audio, and video. False news creators manipulate any or all of these attributes to create false information. Misinformation has undoubtedly become a significant issue affecting businesses, economies, election outcomes, and societies. The dissemination and effects of inaccurate, distorted, or false information spread rapidly and reached millions of people within minutes through various communication channels. False information may be of various forms, such as fake news, disinformation, misinformation, rumors, satire, etc. This book considers all kinds of false information as fake news or misinformation and uses the terms *fake news* and *misinformation* interchangeably. The primary purpose of this book is to provide a foundation for the problems of misinformation or false content detection including various challenges and approaches to solve them. This book is helpful for academicians, researchers, postgraduate as well as undergraduate students, and industry people who wish to explore various dimensions of misinformation detection regardless of their past knowledge and experience.

The book starts with an overall description of misinformation. It briefly introduces the history, various issues or challenges, several reasons for creating and spreading misinformation, several forms, and its impact on individuals and society. The second chapter discusses the prior works on misinformation detection that helped structure the book to solve the various problems. This chapter explores various datasets, recent advancements, and state-of-the-art mechanisms for misinformation detection. The third chapter demonstrates that the presence of surprising content in the story draws instant attention and appeals to strong emotional stimuli in the story reader, which is one of the significant characteristics of the virality of misinformation. It first introduces *novelty, novelty detection, emotion, and emotion recognition*, briefly. After that, it discusses various terminologies that are used throughout the chapter. Subsequent sections explore the application of novelty and emotion in the misinformation detection domain. It also explores a case

study demonstrating a methodology for misinformation detection using emotion recognition and novelty detection. The primary goal of this chapter is to use this concept and implement emotion recognition novelty detection as supporting tasks for automatic misinformation detection. This chapter repurposes textual entailment for novelty detection and classifies disinformation using models trained on extensive entailment and emotion datasets. The fourth chapter explores the importance of the joint learning of interrelated tasks to improve the performance of the primary task. This chapter first introduces multitasking and then discusses its advantages. It focuses on developing a framework for joint learning of interrelated tasks such as emotion recognition, novelty detection, and misinformation detection. More specifically, this chapter justifies that affective information and textual novelty are the main ingredients to consider while designing an automatic misinformation detection framework.

After Chap. 4, this book majorly focuses on the role and importance of multimodality in misinformation detection. The fifth chapter explores various datasets and mechanisms leveraging multimodal information. It first introduces multimodality, its importance, and its applications in various domains for solving different tasks. In continuation, this chapter also talks about challenges in the multimodality. It further discusses a case study for developing a multimodal misinformation detection framework. It aims to leverage various information from different sources (e.g., text, image, video, etc.) to build an efficient system. It justifies that an effective combination of different input modalities may assist the system in proper learning. This chapter also demonstrates the fusion mechanisms of text and image modalities to obtain an efficient multimodal feature that ultimately helps to classify multimedia fake news. The sixth chapter discusses how novelty and emotion can be helpful in multimodal misinformation detection. It justifies that detecting misleading information is difficult without earlier knowledge about that particular news and explores the possible solutions to tackle this problem. Chapter 7 introduces the concept of multilingualism. It starts with discussing multilingual content, its importance, and applications in the NLP domain and continues with the challenges and benefits. It implements an effective neural model to detect fabricated multilingual information, which overcomes the research and development gap in misinformation detection for regional languages as a case study. Finally, the last chapter (Chap. 8) summarizes the contents of the book.

Each chapter starts with an introductory section to briefly introduce the topic. Wherever required, each chapter includes pictorial descriptions of methods and concepts to improve clarity and facilitate better understanding.

Bihar, India — Asif Ekbal
Bhubaneswar, India — Rina Kumari
December, 2023

Suggestions and opinions have always proved very helpful in enhancing any endeavor. We request all readers to write us their valuable feedback/comments/suggestions for the betterment of the book at *asif@iitp.ac.in* or *rina.kumarifcs@kiit.ac.in*.

Acknowledgments

This book is the outcome of a fruitful cooperation between researchers and academicians that has spanned many years. It would not have been possible without the many contributors whose names did not appear on the cover. We thank our institutions (The Indian Institute of Technology Patna (IITP), India, and The Kalinga Institute of Industrial Technology, Bhubaneswar Odisha, India) for giving us the time and opportunity to write the book. We would like to thank all the faculty and staff members of our institutions, who helped us by providing the necessary resources and in various other ways in completing this book. We would like to thank our family, friends, and colleagues for their continuous support and motivation. We also thank the Springer staff for their very professional support and discussions.

Many others not named above contributed to the outcome, and we wish to express our gratitude to all of them.

Contents

About the Authors

Asif Ekbal is currently an Associate Professor in the Department of Computer Science and Engineering, IIT Patna and an Associate Dean, Resources at IIT Patna. He has been pursuing research in the broad areas of artificial intelligence, natural language processing (NLP), and machine learning (ML) since the last 20 years.

He has authored around 300 papers in top-tier conferences and journals. Asif has been involved in several sponsored research projects, funded by industries, such as Elsevier, Accenture, EZDI, LG, Skymap, Samsung Research, Wipro, and Flipkart, and government agencies such as MeitY, MHRD, and SERB, . He has been serving as an area chair, PC member, and reviewer to several well-known conferences. He is currently the associate editor and/or in the editorial board of several reputed journals. He is an awardee of the "Best Innovative Project Award" from the Indian National Academy of Engineering, Govt. of India; "JSPS Invitation Fellowship" from Govt of Japan; and the "Young Faculty Research Fellowship Award" of the Govt. of India. He is among the top 2% of scientists, published by Stanford University findings in Elsevier (2021, 2022, 2023), and top computer scientists, published by Openrsearch.com

Rina Kumari is currently an assistant professor in the School of Computer Engineering, KIIT Deemed to be University, Bhubaneswar, Odisha, India. She holds a PhD from the Indian Institute of Technology Patna, India. Her main area of research is natural language processing, deep learning, and fake news detection. She has published papers in various peer-reviewed conferences like IJCNN and IJCNLP and journals of international repute like *Information Processing and Management* (IPM), *Expert Systems With Applications*, *Journal of Intelligent Information Systems* (JIIS), etc. She is a recipient of the IEEE CIS Participation Grant IJCNN 2021. She has also reviewed several manuscripts of conferences like ICONIP and ICON and journals like IEEE Transactions, Springer, Elsevier (IPM), etc. She has chaired a session in ASSIC 2022.

Acronyms

AMFB	Attention-based multimodal factorized bilinear pooling
ANN	Adversarial neural network
APR	Amazon product review
AUC	Area under the curve
BERT	Bidirectional encoder representations from the transformer
Bi-GRU	Bidirectional gated recurrent unit
Bi-LSTM	Bidirectional long short-term memory
CARN	Cross-modal attention residual network
CNN	Convolutional neural network
CSV	Comma-separated values
DCNN	Deep convolution neural network
DT	Decision tree
EANN	Event adversarial neural network
FNC	Fake news challenge
FND	Fake news detection
FNID	Fake news inference dataset
GCAN	Graph-aware co-attention network
HFFN	Hierarchical feature fusion network
ICMR	Indian Council of Medical Research
IIM	Inter-modal interaction module
KNN	K-nearest neighbor
LARS	Layer-wise adaptive rate scaling
LR	Logistic regression
LSTM	Long short-term memory
MBFC	Media bias/fact check
MCB	Multimodal compact bilinear pooling
MFB	Multimodal factorized bilinear pooling
MKD	Multilingual knowledge distillation
MLB	Multimodal low-rank bilinear pooling
MLP	Multilayer perceptron
MMFND	Multisource multiclass fake news detection

MMM	Multilingual multimodal misinformation
MVAE	Multimodal variational autoencoder
NB	Naive Bayes
NER	Named entity recognition
NLG	Natural language generation
NLI	Natural language inference
NLP	Natural language processing
NPS	Net promoter score
PCA	Principle component analysis
PMTM	Proposed multitask model
QQP	Quora question pair
RBF	Radial basis function
RFC	Random forest classifier
RL	Reinforcement learning
RNN	Recurrent neural network
ROC	Receiver operating characteristics
RTE	Recognizing textual entailment
SCL	Supervised contrastive learning
SGD	Stochastic gradient descent
SOTA	State-of-the-arts
SPO	Subject predicate object
SVM	Support vector machine
TAG	Topic-Agnostic
TER	Textual entailment recognition
TI-CNN	Text and image information-based convolutional neural network
t-SNE	T-distributed stochastic neighbor embedding
USA	United States
VAE	Variational autoencoder
WBG	Without background
WHO	World Health Organization

Chapter 1
Introduction

News and reports contain numerous details, such as author, source, prominent headlines, attractive writing styles, photos, videos, etc. Any variations created to these details result in fake or misleading news. Misinformation or false information is created by manipulating any one or all of these attributes. Misinformation has undoubtedly become a significant issue affecting businesses, economies, election outcomes, and societies. The dissemination and effects of inaccurate, distorted, or false information spread rapidly and reach millions of people within minutes through various communication channels. Fake news aims to emulate the presentation of news and information to obtain legitimacy and credibility. However, fake news is not only limited to gaining the emergence of the news. It seeks to follow journalistic writing, how an authentic news article includes an image, and how a reliable news site presents itself. The quick spread of information has been facilitated by the development of the Internet and the quick uptake of social media sites like Facebook and Twitter. With the help of these platforms, users generate and disseminate more information than ever before, some of it false or misleading and has no bearing on reality. The resemblances of fake news with journalistic news reports amplify the seriousness of this threat regarding the other elements of disinformation. Subsequent sections in this chapter briefly discuss the history of fake news, different definitions, various forms, modalities, clues in fake news, the purpose and reason for creating and spreading misinformation, and the impact of misinformation on society and individuals. It also explores the objective and challenges of misinformation detection.

A. Ekbal, R. Kumari, *Dive into Misinformation Detection*, The Information Retrieval Series 30, https://doi.org/10.1007/978-3-031-54834-5_1