Deepak P • Tanmoy Chakraborty • Cheng Long •
Santhosh Kumar G

Data Science for Fake News

Surveys and Perspectives

Springer

Contents

A Multifaceted Approach to Fake News 1
1 Introduction 1
2 Surveys 3
2.1 Unsupervised Methods for Fake News Detection 5
2.2 Multimodal Fake News Detection 5
2.3 Deep Learning Methods for Fake News Detection 5
2.4 Dynamics of Fake News Diffusion 6
2.5 Neural Text Generation 6
2.6 Fact Checking on Knowledge Graph 7
2.7 Graph Mining Meets Fake News Detection 7
3 Perspectives 8
3.1 Fake News in Health Sciences 9
3.2 Ethical Considerations in Data-Driven Fake News Detection 10
3.3 A Political Science Perspective on Fake News 10
3.4 Fake News and Social Processes 11
3.5 Misinformation and the Indian Election 11
3.6 Science and Technology Studies (STS) and Fake News 12
3.7 Linguistic Approaches to Fake News Detection 12
4 Concluding Remarks 12
References 13

Part I Survey

On Unsupervised Methods for Fake News Detection 17
1 Introduction 17
1.1 Paradigms of Machine Learning vis-à-vis Supervision 18
1.2 Challenges for Unsupervised Learning in Fake News Detection 19

2 Unsupervised Fake News Detection: A Conceptual Analysis 21
2.1 Conceptual Basis for UFND Methods 21
2.2 Critical Analysis of UFND Conceptual Bases 23
2.3 Building Blocks for UFND 26
3 Unsupervised Fake News Detection: A Methodological Analysis 28
3.1 Truth Discovery 28
3.2 Differentiating User Types 30
3.3 Propagandist Patterns 32
3.4 Inter-user Dynamics 33
4 The Road Ahead for Unsupervised Fake News Detection 35
4.1 Specialist Domains and Authoritative Sources 35
4.2 Statistical Data for Fake News Detection 36
4.3 Early Detection 37
4.4 Miscellaneous 38
5 Conclusions 38
References 39

Multi-modal Fake News Detection 41
1 Introduction 41
2 Challenges and Opportunities 44
3 Multi-modal Fake News Datasets 45
3.1 Fake Microblog Datasets 48
3.2 Fake News Datasets 49
4 State-of-the-Art Models 50
5 Unsupervised Approach 51
6 Early Fusion Approaches 54
6.1 JIN 54
6.2 TI-CNN 55
6.3 MKEMN 56
6.4 SpotFake and SpotFake+ 57
6.5 MCE 58
6.6 SAFE 58
7 Late Fusion Approaches 59
7.1 AGARWAL 59
7.2 MVNN 60
8 Hybrid Fusion Approach 62
9 Adversarial Model 63
9.1 SAME 64
10 Autoencoder Model 65
11 Summary of the Chapter 66
References 66

Deep Learning for Fake News Detection 71
1 Introduction 71
1.1 Fake News Types 72
1.2 Early Works 73
2 Deep Learning Methods 74
2.1 Fake News Detection Using CNN 76
2.2 Fake News Detection Using RNN and Its Variants 78
2.3 Multimodal Methods 85
3 Datasets and Evaluation Metrics 87
3.1 Datasets 87
3.2 Evaluation Metrics 89
3.3 Discussion 90
4 Trends in Fake News Detection Using Deep Learning 90
4.1 Geometric Deep Learning 91
4.2 Explainable Fake News Detection 92
4.3 Profiling Fake News Spreaders 93
4.4 Neural Fake News Detection 93
4.5 Discussion 94
5 Conclusion 95
References 95

Dynamics of Fake News Diffusion 101
1 Introduction 101
2 Fake News Diffusion on Facebook 103
3 Fake News Diffusion on Twitter 103
4 Role of Bots in Spreading Fake News 106
5 Trees for Modeling Information Diffusion 107
6 Identifying the Sources of Fake News 107
7 Modeling Fake News Diffusion 109
7.1 Susceptible-Infected-Recovered (SIR) Model 109
7.2 Dynamic Linear Threshold (DLT) 110
7.3 Percolation Model 113
7.4 Spread-Stifle Model 115
8 Strategies to Minimize the Spread of Fake News 123
9 Summary of the Chapter 124
References 126

Neural Language Models for (Fake?) News Generation 129
1 Introduction 129
2 Modeling Approaches 130
2.1 Learning Paradigms for NLG 131
2.2 Language Models 131
2.3 Encoder–Decoder Attention 132
2.4 Autoregression 133
2.5 Seq2Seq Model 133

3 Learning Paradigms 134
3.1 Supervised Learning Techniques 134
3.2 Adversarial Learning Techniques 135
3.3 Reinforcement Learning Techniques 135
3.4 Embedding Techniques 135
4 Pre-trained Language Models 136
4.1 Contextualized Word Vectors (CoVe) 136
4.2 Embeddings from Language Model (ELMo) 137
4.3 BERT 137
4.4 RoBERTa 138
4.5 Transformer-XL 138
4.6 Larger Language Models 139
4.7 XLNet 140
4.8 GROVER 140
4.9 CTRL 141
4.10 Seq2Seq Pre-training Models 142
4.11 Discussion 142
5 (Fake?) News Generation and Future Prospects 143
6 Conclusion 144
References 145

Fact Checking on Knowledge Graphs 149
1 Introduction 149
2 Preliminaries 150
2.1 Knowledge Graph 150
2.2 RDF 151
3 Models 152
4 Knowledge Linker 152
5 PredPath 156
6 Knowledge Stream 162
7 Conclusion and Future Work 166
References 166

Graph Mining Meets Fake News Detection 169
1 Characteristics and Challenges 170
2 Graph Models in Fake News Detection 171
2.1 Information 171
2.2 Graph Models 172
3 Unimodal Scenario: Static Graph-Based Methods 173
3.1 Graph Statistics Detection 173
3.2 Dense Subgraph Mining 176
3.3 Benefits and Issues 178
4 Multi-modal Scenario 179
4.1 Dynamic Graph-Based Approaches 180
4.2 Graph-Assisted Learning Approaches 182
4.3 Benefits and Issues 186

5 Summary of the Chapter 186
References 187

Part II Perspectives

Fake News in Health and Medicine 193
1 Polish Health Misinformation Study 195
2 Stanford University Study: Cannabis, a Cure for Cancer 195
3 NBC News Study 196
4 Dandelion, the Magical Weed 197
5 Polarised Facts 197
6 Fake News During the Pandemic 198
7 Consequences of Health Misinformation 200
8 Managing Health Misinformation 202
References 203

Ethical Considerations in Data-Driven Fake News Detection 205
1 Introduction 205
2 Ethical Dimensions of DFND 206
2.1 Mismatch of Values 208
2.2 Nature of Data-Driven Learning 212
2.3 Domain Properties 215
3 Fairness and DFND 217
4 Democratic Values and Uptake of DFND 222
5 Conclusions 230
References 231

A Political Science Perspective on Fake News 233
1 Introduction 233
2 The Origins of Fake News 236
3 Fake News in the Twenty-First Century 237
4 Fake News and the Study of Politics 239
5 Conclusion 241
References 241

Fake News and Social Processes: A Short Review 245
1 Introduction 245
2 Sociological Studies of Disinformation 248
3 Vaccine Hesitancy 249
4 Elections 251
5 Other Social Processes 253
6 Conclusions 254
References 255

Misinformation and the Indian Election: Case Study 257
1 Misinformation and Disinformation in India 258
1.1 Misinformation and Disinformation in India 258

1.2 Closed Networks for Disinformation 259
1.3 Scale, Prevalence, and Complexity of the Problem 260
1.4 Early Solutions and Fact-Checking in India 261
2 Logically's Capabilities 262
2.1 Automation to Augment Value 262
2.2 Credibility vs. Veracity 263
3 Credibility Assessment 264
3.1 Credibility Assessment 264
3.2 Credibility Assessment Methodology During Indian Elections 267
4 Veracity Assessment 269
4.1 Methodology: The Life Cycle of a Claim 269
4.2 Methodology During Indian Elections 270
5 WhatsApp Solution for a Sharing Nation 273
5.1 Long-Standing Questions 275
5.2 Related Work 275
5.3 Exposing Misinformation on Closed Networks 276
5.4 Disseminating Verifications to Audiences Exposed to Mis/Disinformation 278

STS, Data Science, and Fake News: Questions and Challenges 281
1 Introduction 282
2 Truth, Power, and Knowledge 282
3 Truth Versus Post-truth 284
References 285

Linguistic Approaches to Fake News Detection 287
1 Introduction 287
1.1 Defining Fake News 288
1.2 Linguistics, Sub-disciplines, and Methods 289
1.3 News in Linguistics 291
1.4 Deception in Linguistics 292
1.5 Different Texts and Contexts 292
2 Linguistic Approaches to Fake News Detection 294
2.1 Bag of Words and LIWC 294
2.2 Readability and Punctuation 296
2.3 Deep Syntax 297
2.4 Rhetorical Structure and Discourse Analysis 297
3 Conclusions 297
References 300