

Misbehaviour Detection Algorithms and Application in Social Networks

by Jun Yin

Thesis submitted in fulfilment of the requirements for
the degree of

Doctor of Philosophy

under the supervision of Guandong Xu

University of Technology Sydney
Faculty of Engineering and Information Technology

the September 2021

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, *Jun Yin* declare that this thesis, is submitted in fulfilment of the requirements for the award of *Doctor of Philosophy*, in the *School of Computer Science, Faculty of Engineering and Information Technology* at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Production Note:

SIGNATURE: Signature removed prior to publication.

[Jun Yin]

DATE: 10th September, 2021

DEDICATION

To myself . . .

ACKNOWLEDGMENTS

First of all, I would like to thank Prof. Guandong Xu, my principal supervisor, for his guidance, suggestion, and support throughout my doctoral program course at the University of Technology Sydney. Without his professional guidance and support, this work would not have been achieved.

I am also grateful to my previous co-supervisor Dr. Shaowu Liu for his constructive suggestions on my research work. I would like to pass my gratitude to my current co-supervisor and friend Dr. Qian Li, for her selfless help and suggestions on my work.

In addition, a thank you to Prof. Zhiang Wu of Nanjing Audit University, who introduced me to social network behaviour analysis, and whose enthusiasm for the research work had lasting effect.

Last but not least, I am thankful to my beloved husband, Wei Yuan, and my baby boy Adam J. Yuan, for their love and encouragement which helps me finish this thesis. I am also thankful to my parents and my parents in law for their support throughout my research studies.

Jun Yin
Sydney, Australia
September, 2021

ABSTRACT

The liberty to contribute content freely has encouraged malicious users to exploit the social platforms (i.e., social networks and e-commerce platforms) for their benefits. Spammers, rumours, and some other unexpected activities are almost an appendage to all social platforms that disrupt the network order. We summarize these unexpected activities as misbehaviours in social platforms. To detect such social platform misbehaviours, machine learning is an expected method where modelling and algorithms are two significant elements. Such an interesting topic that has application prospects and research value has attracted the attention of many researchers, and some results have also been put forward in the literature.

In terms of spammer detection, because of the rich data types of e-commerce platform, such as score, comment content, and comment time, the mainstream detection methods rely on the above data to construct features on e-commerce platforms. For example, text-based features, behaviour features etc. in conjunction with some supervised learning algorithms like Naive Bayes, Decision trees etc. are the most frequently used combinations for spammer detection in e-commerce platforms. However, social networks are based on interaction data and are relatively deficient in data types, thus, spammer detection in social networks requires a detection framework that relies on relational data but is independent of content data. Along this line, the existing research attempts to define complex network features (e.g., degree, K-Core, PageRank, connected component, etc.) and interactive sequence-based features. Nevertheless, the deep semantic information hidden in the multi-relational networks has not been fully utilized.

Furthermore, rumour as another type of misbehaviours in social networks has been run through the whole evolutionary history of mankind. People maliciously disseminate rumours to increase awareness, slander others or cause panic, etc. To eliminate this issue, many researchers resort to detecting rumours in social networks. However, rumour detection is not sufficient to eliminate the negative impact, which also requires official institutions to provide the refutations. In practice, the number of rumours in social networks is too large, there is no need to refute some rumours with little or no concern. Therefore, an evaluation of the impact of the rumours in advance is essential.

To address the aforementioned research problems, a few approaches are proposed in the works introduced in this thesis.

- Based on the non-content data, we fully excavate the deep semantic information hidden in the heterogeneous network and define a series of user behaviour features using relational network data for spammer detection.

-
- Based on the graph embedding method, we propose a “Send-Receive” Role Separable Graph-Embedding Model (*RS-GEM*) to extract and fuse the hidden features of heterogeneous relations in multi-relational social networks to detect spammers.
 - Inspired by deep sequential networks, we propose a “Multi-level Dependency Model” (*MDM*), which exploits user’s behaviours in terms of long-term and short-term dependency from both individual-level and union-level to detect multi-relational social spammers.
 - Before a rumour has an impact on social networks, we need to assess the possible impact it may have. Therefore, we devise a rumour influence prediction model *RISM* (Rumour Impact on Social Media) based on a popular rumour intensity formula to predict the impact of a newborn rumour.

Last but not least, since the global outbreak of COVID-19 in early 2019, COVID-19-related topics have become hot spots on social networking platforms. At the end of this thesis, we shall also analyze the COVID-19-related tweets on Twitter and get a preliminary understanding of the public’s focus and sentiment trends during the pandemic.

Keywords: Multi-relational Social Network; Behaviour Analysis; Spammer Detection; Feature Construction; Rumour Analysis; Sentiment Analysis

LIST OF PUBLICATIONS

RELATED TO THE THESIS :

1. Yin, J., Zhou, Z., Liu, S., Wu, Z., & Xu, G. (2018, June). Social Spammer Detection: A Multi-Relational Embedding Approach. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 615-627). Springer, Cham.
2. Yin, J., Liu, S., Li, Q., & Xu, G. (2019, October). Prediction and Analysis of Rumour's Impact on Social Media. In 2019 International Conference on Behavioral, Economic, Socio-cultural Computing (BESC) (pp. 1-6). IEEE.
3. Yin, J., Li, Q., Liu, S., Wu, Z., & Xu, G. (2020, November). Leveraging Multi-level Dependency of Relational Sequences for Social Spammer Detection. *Neurocomputing*, vol. 428, pp. 130-141.
4. Yin, J., Li, Q., Liu, S., & Xu, G. Social Network Analysis of Twitter User's Behaviour during COVID-19 Pandemic. Prepared to be submitted as a Journal Paper.

OTHERS :

5. Zhou, Z., Liu S., Xu, G., Xie, X., Yin, J., Li, Y., & Zhang, W. (2018, June). Knowledge-Based Recommendation with Hierarchical Collaborative Embedding. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 222-234). Springer, Cham.

TABLE OF CONTENTS

List of Publications	ix
List of Figures	xv
List of Tables	xvii
1 Introduction	1
1.1 Background	1
1.2 Research Objectives	3
1.3 Proposed Approaches	4
1.3.1 Proposed Approaches for Spammer Detection	5
1.3.2 Proposed Approaches for Rumour Analysis	6
1.3.3 Proposed Approaches for COVID-19 Related Analysis	7
1.4 Thesis Outlines	7
2 Literature Review	11
2.1 Spammer Detection	11
2.1.1 Spammer Detection in E-commerce Platforms	13
2.1.2 Spammer Detection in Social Networks	15
2.2 Rumour Classification Process	18
2.3 Rumour Detection	20
2.4 Summary	21
I Part I Spammer Detection	23
3 Data Description and User Behaviour Analysis	25
3.1 Data Source and Scale	26
3.2 Social Relations Speculation	26

TABLE OF CONTENTS

3.3	User Behaviour Analysis	33
3.3.1	Misbehaviour Hypothesis and Feature Construction	34
3.3.2	Effectiveness Analysis	37
3.4	Summary	41
4	Graph-embedding based Social Spammer Detection	43
4.1	“Send-Receive” Role Separable Graph-Embedding Model (<i>RS-GEM</i>)	44
4.1.1	Motivation	44
4.1.2	Framework Overview	46
4.1.3	Model Formulation	47
4.2	Experiment and Analysis	49
4.2.1	Experiment Setup	49
4.2.2	Effectiveness Analysis	50
4.2.3	Cross Validation and Comparison	53
4.3	Summary	55
5	Sequence-based Social Spammer Detection	57
5.1	Multi-level Dependency Model	58
5.1.1	Motivation	59
5.1.2	Framework Overview	61
5.1.3	Model Formulation	64
5.2	Experiment and Analysis	69
5.2.1	Experiment Setup	70
5.2.2	Experimental Results	72
5.2.3	Discussion	75
5.3	Summary	76
II	Part II Rumour Analysis	77
6	Rumour Impact Analysis on Social Media	79
6.1	Introduction	79
6.2	Preliminaries	81
6.2.1	Problem Statement	81
6.2.2	Measuring Rumour Impact	81
6.3	Methodology	83
6.3.1	Importance	83

6.3.2	Ambiguity	83
6.3.3	Public Critical Ability	85
6.3.4	Content-based Feature Extraction	86
6.4	Experiment and Analysis	88
6.4.1	Dataset	88
6.4.2	Experiment Metrics	89
6.4.3	Comparison	89
6.5	Summary	90
7	Sentiment Analysis towards COVID-19 on Twitter	93
7.1	Introduction	93
7.2	Preliminaries	95
7.2.1	Data Source and Scale	95
7.2.2	Initial Analysis	99
7.3	Topic Modeling and Sentiment Analysis	101
7.4	Discussion	103
7.5	Summary	105
III	Part III Conclusion	107
8	Conclusion and Future Work	109
8.1	Contributions	109
8.2	Possible Future Work	111
	Bibliography	113

LIST OF FIGURES

FIGURE	Page
2.1 Processes of rumour classification	19
3.1 Relations speculation.	29
3.2 Illustration of relations speculation using elimination.	32
3.3 CDF curve of feature FOTD.	38
3.4 CDF curves of features RSR & RU.	39
3.5 CDF curve of feature RSST.	39
3.6 CDF curves of features EXT & RA.	40
4.1 Motivation of RS-GEM. The middle user sends messages to too many users, which makes him suspicious, like a spammer. However, he received gifts from users i and j , which is a powerful indicator that he is a good user. But we found that the users who gave the gift are actually low-credit users who were blocked by others, so the fact may be that the spammer is trying to deceive the detection system.	45
4.2 Illustration of the RS-GEM on two relations.	47
4.3 Effectiveness analysis of latent features.	51
4.4 Effectiveness analysis of latent features (continued).	52
4.5 Impact of the number of dimensions in our <i>RS-GEM</i> model.	55
5.1 Examples of individual-level dependency among the relational sequence. Normal users may be involved in one of the sequences of relations given in (a), (e.g., <i>add friend</i> first, and then send <i>message</i>). On the other hand, spammers may only imitate one or two relational behaviours of normal users (e.g., <i>add friend</i> first, then send <i>message</i>). However, for the reason that spammers always have their own malicious purposes, thus, they cannot completely imitate all the behavioural sequences of a normal user in (a).	60

LIST OF FIGURES

5.2	Examples of union-level dependency among the relational sequence. Union-level dependencies can capture the collective impact between relational unions performed by users to some extent. For instance, the normal user is more likely to view profile, add friend, give a gift and sending messages together than view profile, add friend, give a gift or send messages individually. This combination of behavioural sequences makes it much more difficult for spammers to imitate the normal users.	61
5.3	Framework of Multi-level Dependency Model (<i>MDM</i>).	63
5.4	Multi-order Attention Network.	67
5.5	Performances of MDM under different sequence lengths n	74
5.6	Performances of MDM under different embedding sizes d	74
6.1	Framework of RISM.	84
7.1	Number of COVID-19 related tweets from February 1, 2020 to May 31, 2020.	96
7.2	Word clouds of each month from February 1, 2020 to May 31, 2020. . .	100
7.3	The frequency distribution of tweets on five main topics across sentiment types from February 1, 2020 to May 31, 2020	102
7.4	The frequency distribution of tweets on five main topics across sentiment types on each month from February 1, 2020 to May 31, 2020.	103

LIST OF TABLES

TABLE	Page
2.1 Representative features for reviews on e-commerce platforms	15
3.1 Attribute description of <i>Tagged.com</i> user identity information dataset	27
3.2 Attribute description of <i>Tagged.com</i> users' interaction records dataset	27
3.3 Statistics of 7 relations within 1 day	28
3.4 Statistics of edges of 7 relations within 1 day	30
3.5 Statistics of the number of <i>src</i> user's identity (spammer / normal user)	31
3.6 Statistics of the number of <i>src</i> user's (spammer / normal user) interaction records	31
3.7 Relation ID and corresponding semantic relation type name	33
3.8 Behavioural features based on Hypothesis 1	35
3.9 Behavioural features based on Hypothesis 2	36
4.1 Statistics of experiment dataset	50
4.2 Comparison of two classifiers with different kinds of features	54
5.1 Summary of notations	64
5.2 Statistics of <i>Tagged.com</i> dataset	71
5.3 Performance comparison with baselines (the best result of each metric is bold)	73
5.4 Performance comparison on different components in MDM (+ represents adding a layer to the last row, and the best result of each metric is bold)	75
6.1 Statistics of Corpus	88
6.2 Statistics of Impact Scores	88
6.3 Statistics of Dataset	89
6.4 Comparison of Four Classifiers with Different Kinds of Features	90
7.1 Actively tracking keywords in Twitter according to COVID-19	97

LIST OF TABLES

7.2	Actively tracking accounts in Twitter according to COVID-19	97
7.3	Attribute description from <i>Hydrator</i> for each tweet	98
7.4	Number of COVID-19 related tweets in each month after screening	99