Mapping Information Diffusion through Social Computing for Knowledge Discovery

Thesis submitted to
Cochin University of Science and Technology
in partial fulfilment of the requirements for the award of the degree of

DOCTOR OF PHILOSOPHY

by
MINI U

Under the guidance of

Dr. K. POULOSE JACOB

FACULTY OF TECHNOLOGY
Cochin University of Science and Technology
Kochi - 682 022, Kerala, India

January 2018

Mapping Information Diffusion through Social Computing for Knowledge Discovery

Ph.D. Thesis in the field of Social Media Mining

Author

MINI U
Department of Computer Science
Cochin University of Science and Technology
Kochi– 682 022, India
E-mail: mini_u@cusat.ac.in

Research Advisor

Dr. K. POULOSE JACOB (Supervising Guide)
Professor, Department of Computer Science
Cochin University of Science and Technology
Kochi– 682 022, India
E-mail: kpj@cusat.ac.in
JANUARY 2018

Dedicated to …

My Family, Friends & Well wishers

DECLARATION

I hereby declare that the thesis entitled " ***Mapping Information Diffusion through Social Computing for Knowledge Discovery***" is the authentic record of research work carried out by me, for my Doctoral Degree under the supervision and guidance of Dr. Poulose Jacob K Professor, Department of Computer Science, Cochin University of Science & Technology and that no part thereof has previously formed the basis for the award of any degree or diploma or any other similar titles or recognition.

KOCHI- 22 Mini U

30-01-2018

FACULTY OF TECHNOLOGY
COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY
KOCHI- 682022, INDIA

Certificate

*Certified that the work presented in this thesis entitled "**Mapping Information Diffusion through Social Computing for Knowledge Discovery**" is based on the bona fide research work done by Ms. Mini U under my guidance in the Department of Computer Science, Cochin University of Science and Technology, Kochi -22 and has not been included in any other thesis submitted previously for the award of any degree.*

Kochi-22
30/01/2018

Dr. K. Poulose Jacob

Certificate

This is to certify that all the relevant corrections and modifications suggested by the audience during the Pre-synopsis seminar and recommended by the Doctoral Committee of the candidate have been incorporated in the thesis.

Kochi-22
30/01/2018

Dr. K. Poulose Jacob

Acknowledgements

The research work leading to PhD has been a long and challenging journey. This work would not have been completed, but for the unstinted support and guidance of many. I thank God Almighty for the blessings which have led to the completion of this research work.

I express my sincere gratitude to my guide Dr. K.Poulose Jacob, Professor, Department of Computer Science, Cochin University of Science and Technology, for his encouragement, guidance and the immense support he extended to me through out. I am privileged to have him as my guide.

I thank Dr. Sumam Mary Idicula, Professor, Department of Computer Science, Cochin University of Science and Technology, for motivating me in pursuing the PhD programme in the department. I am grateful for all her suggestions in my learning process. I am extremely thankful to Mr.K.B.Muraleedharan and Dr.Jacob Philip for the support extended for the completion of my course work.

I had the support and guidance of Dr. Kannan and Dr. Judy, Department of Computer Applications from whom I got constant inspiration to work hard

and to be persistent. Their insightful comments, support and advice has been of immense benefit to me during the most difficult times. They always kept me motivated and inspired during this entire journey.

My sincere thanks is due to Dr. G Santhosh Kumar, Professor, Department of Computer Science, Cochin University of Science and Technology, for all the help he has extended. I express my sincere gratitude to him. I am equally thankful to Dr. Baby.M.D, Dr.M.Bhasi, Dr. K.V. Pramod, Dr. Santosh Kumar M.B, research scholars of Department of Computer Applications and Department of Computer Science for their constant support and motivation. I would like to thank Ms. Liny Varghese who helped me with ideas and final conceptualisation of the thesis.

I am thankful to Mr. Vijay Paul in sharing the data sets during my initial period of research which was extremely helpful in pursuing with the research problem. I am indebted to Mr. Vijay Nair for his data sharing and insightful interventions which helped me to complete the research. I am also thankful to Ms. Delna Gomez for the layout of my thesis.

I have special word of appreciation for Mr. Joe Joseph in providing me the best library support. I also place on record my gratitude to Mr. Lal Paul, Mr. Renjith and Mrs. Manju for providing me all the technical support required for carrying out my research work. I am extremely grateful to Ms.

Girija and all the staff of the department for their encouragement and support. A special gratitude and appreciation is extended to my colleagues and professional friends. My friends are my life. I thank all my friends for the support they had given me in difficult moments. I am deeply indebted to them for this.

My husband Harikrishnan, daughter Keerthana and my father Rajagopal showered unconditional support and encouraged me to pursue my dreams. They have been my source of strength. I could never have reached this point without their encouragement and support. I express my profound gratitude to all who helped me in this journey.

MINI U

Abstract

Computational Social Science is an emerging discipline that uses digital tools to analyze the rich and interactive life we lead. The exponential growth of Internet technologies, World Wide Web and social media in the last few years are generating large amount of social and behavioural data through the social interactions. The data about human behaviour is available for further analysis on a scale never thought about and with tremendous granularity as well as precision. This is helping social scientists to study human behaviour and social interaction in unprecedented detail. These techno-sociological studies need collaborative, interdisciplinary contribution of computer experts, information scientists, physicists, as well as mathematicians and sociologists. The ability to collect and process social data helps researchers to address core questions in social sciences in new ways which opened up nascent areas to explore. Talking to friends in our social network, navigating the Web and forming opinions by listening to others and to the media have become part of daily life. Challenges have arisen recently with the advent of online social media, which produces large amounts of both network and natural language data. Thus understanding, predicting, and enhancing human behaviour in networks pose important research problems for computer and data scientist with practical applications of high impact.

Systems as diverse as genetic networks or the World Wide Web are best described as networks with complex topology.

Computational Social Science uses powerful computer simulations of networks, data collected from online social networks and experiments involving hundreds of thousands of individual conversations to answer questions that were previously impossible to investigate. Networks have been studied as graphs in mathematics, physics, sociology, engineering and computer science, biology and economics. Grounded in graph and system theories, this approach has proven to be a powerful tool for studying networks in physical and social world, including the web. Social media has become an integral part of any business promotions today, unlike the earlier times where it was only considered only as a social networking tool. This growing pace of social media and its impact has evolved in such a way that it is continuously shifting, leaving marketers constantly challenged, and most businesses overwhelmed with the never-ending changes. Social media analysis has become a necessity for business. The challenge for marketers is to find new ways to capture the attention of consumers who are bombarded with too much digital noise or information every day.

Objectives of the Research Work

The objective of the research is to understand the information diffusion process in an online social network, develop a procedure using text mining techniques to do opinion mining, listen to the social media conversations and extract business intelligence. The research is undertaken in three parts:

- Study the application of social network analysis and understand various parameters
- Gain knowledge on how the information is diffused in a social network.
- Investigate the social customer relationship management with a view to evolving a knowledge discovery process leading to business intelligence.

Information and Activity Diffusion and Propagation: The propagation of information and activities through a social network is an area less investigated or studied from a research perspective specifically to India. India being the second largest in the number of Internet users, the business organisations have great interest in such studies to understand the spread of information which may be applied in promoting their products with a view to achieving their goals.

The user interactions are tracked using Facebook Insights. It helps in tracking the number of active users which leads to understanding page performance. Various metrics are analysed to measure the effectiveness of the social media campaign. The social media mining in viral marketing is studied on current and potential adopters. This is compared against the Bass Model, which is the standard diffusion model in order to describe the process of how new products get adopted in the market.

Crowd sourcing and Opinion mining in Social Media: - Customer sentiment analysis is a method of processing information, generally in text format, often from social media sources, to determine customer opinion and responses. Analysis of the data allows organizations to assess whether customer reaction to a new product was positive or negative, or whether owners of a product are experiencing major technical difficulties. Social Customer Relationship Management (CRM) can be integrated with the website to use social networking portals for furthering the business using viral marketing. This is a two sided weapon which can promote your product via a viral positive feedback, or do the opposite by a similar negative one. Hence promoting bi-directional communication between companies and customers using social media and analysing the customer feedback to use crowd sourcing for business benefit is the need of the hour. Analysis of aggregated data over time provides insights into trends, while analysis of individual cases in near real time lets companies address and

resolve customer issues quickly. At the heart of customer sentiment is text analysis, a complex process based on statistical and linguistic analyses.

A tool **SENTIMATCH** is developed to understand opinion, emotion and sentiment in contextual level beyond keyword searching. The domain selected is eco-tourism. This tool is integrated into a portal which maps all the conversation and this promote and monitor public participation. Sentiment analysis is done on these conversations which gives a good snapshot of what the relationships within and between groups in a country are, at a specific point in time. The methods used here can be extended for any products with public opinion.

Applicability to Social Intelligence: Organisations have begun to use social media, to enable participation and knowledge sharing a relatively recent phenomenon, with the aim of improving business operations. Knowledge Management is a process of blending the internal and external information of an organisation to acquire an actionable knowledge using the different forms of technological platform. Social media can support a range of knowledge management (KM) practices. Despite a number of researchers recognizing the importance of social media to KM, there are currently few studies reported. New ways of social interaction through computer systems are transferable from the general context of the Internet to corporate intranets, where they provide support to Knowledge Management.

The framework realises the theory presented in the work. The focus is on possible applications of the automated sentiment analysis and how the framework can be helpful in providing insightful views.

CONTENTS

CHAPTER -II

CHAPTER -III

List of Tables

List of Figures

List of Abbreviations

SNA	Social Network Analysis
OSN	Online Social Networks
SNM	Social Network Mapping
SCC	Strongly connected component
WCC	Weakly connected component
UGC	User Generated Contents
CRM	Customer Relationship Management
KM	Knowledge Management
WOM	Word of Mouth
NER	Named Entity Recognition
KD	Knowledge Discovery
BI	Business Intelligence
UIMA	Unstructured Information Management Applications

CAQDAS	Computer Assisted Qualitative Data Analysis Tools
JDD	Joint Degree Distribution
RST	Rhetorical Structure Theory
PMI	point wise mutual information
SVM	Support vector machine
NB	Naive Bayes
NLP	Natural language processing
IE	Information Extraction
RDF	Resource Description Framework
OWL	Web Ontology Language
TF- IDF	Term frequency – inverse document frequency
URI	Uniform Resource Identifier

FB	Facebook
BM	Bass Model
ROI	Return on Investment
G/SG	Gamma/shifted Gompertz distribution
UIMA	Unstructured Information Management Architecture
NLTK	Natural Language ToolKit
ETL	Extract, Transform and Load
VADER	Valence Aware Dictionary and sentiment Reasoner
POS	Part Of Speech
NER	Named Entity Recognizer
CRM	Customer relationship management
SKOS	Simple Knowledge Organization System

RSS	Rich Site Summary
CSV	Comma Separated Value