Opinion Influence Modeling in Social Media

Chen Chengyao

PhD

The Hong Kong Polytechnic University

2018

The Hong Kong Polytechnic University

Department of Computing

Opinion Influence Modeling in Social Media

Chen Chengyao

A thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

June 2018

Certificate of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

____________________________(Signed)

Chen Chengyao (Name of student)

Abstract

Online social media have gained a lot of popularity and experienced a fast growth in the past decade. The emergence of social media offers ordinary persons remarkable opportunities to create messages expressing their opinions. Besides, by establishing relationships with others, people can easily convey their opinions to others. Opinion influence produced by social interaction becomes an important factor for people to adapt their behaviors and make decisions. Understanding opinion influence would greatly benefit a variety of marketing activities, such as spread of ideas, public opinion monitoring, and intervention. This thesis aims to provide insights into opinion influence modeling from three components: user interaction, temporal dynamics and the exchanged content. Twitter datasets, which contain the users' opinion traces and user network structures, are collected for the study of opinion influence.

The first work investigates the temporal properties of opinion behaviors. Opinion influence is produced by long-term interactions, where a user continuously collects opinions from neighbors and further changes her/his own opinions accordingly. The temporal dynamics is of great importance for uncovering the underlying mechanism of opinion influence, but is ignored in the current research work. Here, we propose a temporal opinion influence model, which is able to track the opinion dynamics of each individual user and uncover opinion influence through correlating opinion dynamics of connected users. Specifically, we propose two indicators to capture the effects of social interaction on the opinion formation, including friend effect and opinion effect.

The second work delves into the textual content exchanged during communication. The textual message, as the medium of social interaction, provides the foundation to understand communications between users. Rooted in neural network technology, a content-based opinion influence model is proposed to study how opinion influence is driven by the content. Apart from the exchanged content, we also consider the identities of users involved in communication. Each user is characterized by a personal identity and a social identity. A joint learning framework is developed to detect the social identities of users and models opinion influence concerning different user identities at the same time. This work first goes a step further to introduce the content into opinion influence modeling. Its novel idea of integrating personal images in the understanding of user opinion influence also contributes.

In the third work, all the components explored in the above two studies, including the temporal dynamics of user interactions and the content included in the exchanged texts, are carefully considered. Inspired by the advances of the recurrent neural network in sequence modeling, a sequential content-based opinion influence model is developed. It offers to predict opinion words other than opinion sentiment, which may benefit marketing analysis in a more comprehensive manner. This work provides a complete and effective understanding of the opinion influence process. It can be further extended to model the content-based user dynamics in other scenarios.

We conduct a systematical study to understand opinion influence on social media from three components. Our study benefits a variety of marketing activities, including advertisement dissemination, optimization of product impact and other business intelligence related applications. Besides, other complex dynamics of human behaviors, from buying behaviors in the business to voting behaviors in the politics can be unrevealed by continued extension of the proposed framework.

Publications

1. Ziqiang Cao, **Chengyao Chen**, Wenjie Li et al. TGSum: Build Tweet Guided Multi-Document Summarization Dataset. In AAAI. 2016.

2. **Chengyao Chen**, Dehong Gao, Wenjie Li, and Yuexian Hou. Inferring topic-dependent influence roles of twitter users. In Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval, pages 1203-1206. 2014.

3. **Chengyao Chen**, Dehong Gao, Wenjie Li, and Yuexian Hou. A constrained multi-view clustering approach to influence role detection. In Proceedings of the 1st International Conference on Social Influence Analysis Workshop @IJCAI. 2015.

4. **Chengyao Chen**, Zhitao Wang, Yu Lei and Wenjie Li. Content-based influence modeling for opinion behavior prediction. In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics. 2016.

5. **Chengyao Chen**, Wenjie Li, Dehong Gao, and Yuexian Hou. Exploring interpersonal influence by tracking user dynamic interactions. IEEE Intelligent Systems. 2017.

6. **Chengyao Chen**, Zhitao Wang, and Wenjie Li. Modeling opinion influence with user dual identity. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. 2017.

7. **Chengyao Chen**, Zhitao Wang, Wenjie Li and Xu Sun. Modeling Scientific Influence for Research Trending Topic Prediction. In AAAI. 2018.

8. **Chengyao Chen**, Zhitao Wang, and Wenjie Li. Tracking dynamics of opinion behaviors with a content-based sequential opinion influence model. IEEE Transactions on Affective Computing. 2018.

9. Yu Lei, Qiang Chen, **Chengyao Chen** and Wenjie Li. User Preference Modeling by Trust Propagation for Rating Prediction. In IEEE International Conference on Smart City. 2015.

10. Zhitao Wang, **Chengyao Chen** and Wenjie Li. Predictive network representation learning for link prediction. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2017.

11. Zhitao Wang, **Chengyao Chen** and Wenjie Li. Attention Network for Information Diffusion Prediction. In Companion of the The Web Conference. 2018.

Acknowledgements

First and foremost I want to thank my supervisor Prof. Wenjie Li for her guidance, continuous support and great advice during my Ph.D. study. On the academic level, Maggie taught me a lot in how to become a qualified researcher in the computer science area. Under her supervision, I learned how to define a research problem, find a solution to it, and finally present the results to others. On a personal level, Maggie inspired me by her hardworking and passionate attitude. It is my honor to be her student, and I could not imagine how I can survive the Ph.D. study without her help.

I want to thank my kind and talented lab mates that include Dehong Gao, Yanran Li, Zhitao Wang, Ziqiang Cao, Yu Lei and etc for their continued support. I am lucky to join Dr.Li's group, and this dissertation would not have been possible without their constructive suggestions and help. I would also like to thank my dear friends Chunxiao Zhang, Muyan Li, Wei Lu, Jia Hou for their motivation, inspirational energy, and strong support.

Last but not the least, I dedicate this dissertation to my family for all their constant and unconditional love. I would like to express my deepest gratitude to my parents, who raised me with all their love and care and gave me a lot of freedom to pursue my own interests. And most of all, I would like to thank my loving, encouraging and supportive husband, Chongyang Li. He gave my faithful support during this hard journey. Without him, I cannot achieve what I achieve today.

Table of Contents

List of Figures

List of Tables

List of Notations

Notation	**Description**
N	the number of users .
V	a set of $\lvert V\rvert = N$ users.
n_u	the number of u's posting messages
m_u	the number of u's neighbors
F_u	a set of u's neighbors in the network.
$S_u = \{M_u(1), \cdots, M_u(n_u)\}$	u's posting records.
$M_u(i) =< W_u(i), o_u(i), t_u(i) >$	triples representing the tweet information
$W_u(i)$	opinion words included in u's i-th message
$o_u(i)$	the sentiment of u's i-th message.
$t_u(i)$	the time u publishes the i-th message.

Chapter 1

Introduction

1.1 Background

Social media connects users in an exciting way and affects almost all the facets of our daily life. Twitter, a microblogging website, allows people to share information and to interact with each other. Epinion, a review website, provides people with an online platform to share their purchasing experiences. Quora, an online community Q &A platform, offers people opportunities to ask and answer questions and to vote for the answers they agree best. Social media occupies a large part of everyone's life. In 2017, 81 percent of Americans had at least one social media account and their time spent on social media was 135 minutes per day in average[1]. Not only the way people communicate has been changed by the popularity of social media, but also the way people run their business. Hereby, understanding human behaviors on social media is of vital importance for a variety of parties, ranging from marketers to politicians [163].

One notable feature of social media different from traditional media is that it provides people with opportunities to create web content in textual, visual or acoustic types, and it revolutionizes the way how information is created [86]. In the past, public information is only created by the media professional such as TV, radio,

[1]https://www.statista.com/markets/424/topic/540/social-media-user-generated-content/

newspapers and so on. At that time, it is impossible for the opinions of an ordinary person to reach out to the public. Nowadays, social media has become an important channel for an ordinary person to express her/his opinions on all facets of life. It is common to see that a customer posts her/his reviews on Yelp about the delayed service while sitting in a cafe, or a user posts a tweet on Twitter complaining that her/his iPhone bends in the pocket. Consumers no longer merely act as passive recipients of product-, brand- or firm-related information, but are enabled to create, modify and exchange their personalized content about the products through social media [86]. In the meanwhile, the user-generated content (UGC) provides a direct way for enterprises to understand their customers' suggestions and demands, which is valuable for product development and service improvement in future [38].

In addition to providing a channel for people to publish personal opinions, social media like Twitter also brings interactivity to people, which makes users directly communicate with one other without the limitation of time and location [99]. People could establish relationships with anyone they are interested in and initiate conversations with different people on any topic concerning their common interest. These interactions and conversations via social media become important factors for people to adapt their behaviors, revise their judgments and make decisions [120], especially decisions on a business product [84]. In fact, in a marketing study, 71% of consumers said they are more likely to make a purchase based on social media referrals[2]. The opinion influence by which people change their minds as a result of social interactions plays a prominent role in many marketing activities, such as the spread of ideas, public opinion monitoring, and intervention [121]. Given the remarkably large scope of marketing activities that are shaped by opinion influence, it is necessary to explore the mechanisms behind the opinion influence process.

[2]http://www.socialmediatoday.com/content/30-statistics-how-social-media-influence-purchasing-decisions-infographic

With the increasing variety and volume of social media data, computer science and business are becoming more and more intertwined. Social media generates various types of data related to users' profiles, such as their opinion dynamics and their social relationships every day. The large amount of information generated by social media is too difficult for manual analysis. It is necessary to create the computational methods that can process and respond to all of the information. Opinion influence modeling, which is of critical concern to marketers, is a representative problem where both the knowledge from marketing theories and computational approaches from computer science are needed. The target of this thesis is to ***find practical solutions to the problem of opinion influence modeling with an application for opinion prediction***. What is the underlying mechanism of opinion influence process? How to measure the interpersonal influence between users quantitatively? How do people adjust their opinions in the social interactions?

1.2 Research Motivation

Sentiment classification of opinions and influencer detection on social media are two attractive topics in current research areas that are relevant to this thesis. Unfortunately, existing work in both areas are inadequate for understanding and estimating opinion influence. Sentiment classification methods detect the sentiment of a posting as positive or negative. The content information included in the message is explored, yet the process of opinion exchange between users is not captured. Thus, the sentiment classification methods do not have the ability to uncover the opinion influence process produced by the social interactions. In the research study of influencer detection, the network features and topical similarities were both considered to measure the influence powers of users within a network. However, the macro-level influence only represents the overall influence a user exerts on all other users, but cannot account for the influence one exerts on individual followers. In the line of study on information diffusion, researchers studied the diffusion probability between each pair of users, which characterized the chance a message would be propagated from one person to another. However, the proposed diffusion models mainly focus on one-time activation behavior, such as the retweeting behavior of a specific message. They cannot cope with the temporal behaviors that users continuously change their opinions when collecting information from their friends. All these problems pose both challenges and opportunities to the research in this thesis.

We argue that the opinion influence occurs when an individual's opinions are affected/changed by other people. Different from previous work, we study the opinion influence from an interpersonal view and define **opinion influence** in a more systematic way as a framework of "***who influences whom on what opinion/feeling***". To better capture the interpersonal opinion influence, we stress on monitoring users' opinion changes during a long time period. In reality, people are constantly exposed to a flow of opinions created by their following neighbors. Every time after receiving the opinions from her/his neighbors, an individual tends to filter and integrate the opinion influence from the neighbors and adjust her/his own beliefs accordingly [178]. During long-term communication and interaction, the repeated opinion influence would result in opinion dynamics, *i.e., a user continuously changes her/his opinions towards a given issue.*

The problem of opinion influence has been theoretically studied by researchers from a viewpoint of psychology since several decades ago. In most of their studies, the opinion was referred as the belief or judgment of a person towards an issue, such as the emotion expressed in the picture or the sentiment polarity towards a product. According to the observations from the in-house laboratory experiments, researchers formulated opinion influence as an aggregation process and developed different assumptions to describe the aggregation process about how a user formed a new opinion. It can be averaging over all neighbors' opinions [46], following the majority opinion [40], or more sophisticated formulations [43]. Though these studies provide informative understandings about the opinion influence process from a variety of viewpoints, they share significant drawbacks. They limit opinion influence in the setting of laboratory experiments, which simplifies user interactions in real situations. Besides, their assumptions are only verified through computer simulation but lack empirical verification on real datasets. Despite all these efforts, a systematic and practical study for opinion influence modeling on social media and an empirical evaluation on the estimated opinion influence, are needed.

In this thesis, we concentrate our studies on: (1) verifying the existence of opinion influence on social media (2) developing various models that can "learn" interpersonal opinion influence from existing communication records; (3) uncovering the interaction-driven opinion influence process by exploring content of opinions and monitoring temporal opinion dynamics; and (4) forecasting people's future opinions by exploiting estimated opinion influence. The developed framework has many potential applications, such as inference of potential opinion propagation and public opinion intervention.

1.3 Research Overview and Contributions

Psychologists have already designed numerous laboratory experiments to study the mechanism of how a user adapts to the new opinion after receiving the advice from other people. Though the sophisticated laboratory experiments have been designed by researchers, the artificial environment of the laboratory is still so far from real-life, and the findings gained from these experiments could tell little about people's behaviors in real life. With the advent of social media, we could easily observe the online communication among people. It provides researchers opportunities to analyze human behaviors using their online traces. Considering differences in properties between online communication and offline communication, we start the study with a statistical test to verify the existence of the opinion influence by using Twitter datasets related to commercial products. A shuffle test approach is proposed in Chapter 3 and the statistical results demonstrate that a user actually revises her/his opinions towards a specific product after receiving opinions from her/his friends. Then a systematic study is conducted to quantitatively measure and evaluate opinion influence. The study of the opinion influence process incorporates three critical components: user interaction, temporal dynamic, and opinion content. Figure 1.1 presents a sketch of the opinion influence process. Opinion influence appears when people interact with each other and exchange their opinions. It is necessary to understand the opinion influence from the interaction point of view. During the long-term interaction, an individual's opinion may change by being repeatedly influenced by her/his neighbors, which produces the temporal opinion dynamics. Modeling the opinion influence process requires a deep understanding of user interactions and their temporal properties. Besides, opinion is usually expressed through the textual content on social media. The different styles of expression produce different effects on people. This thesis focuses on the above three critical components and spends three chapters to discuss their effects in opinion influence modeling.

In work 1, we investigate temporal properties of opinion behaviors, and model opinion influence by uncovering interaction effects between users. We then focus on exploration of user-generated content, especially the textual messages exchanged in the opinion influence process in work 2. In work 3, temporal property and content information are both integrated into a unified framework, which delivers a comprehensive view to understand opinion influence. Each work is introduced in one chapter (from Chapter 4 to Chapter 6). With the learned opinion influence model, future opinion sentiments of all network users can be predicted, and opinion influence

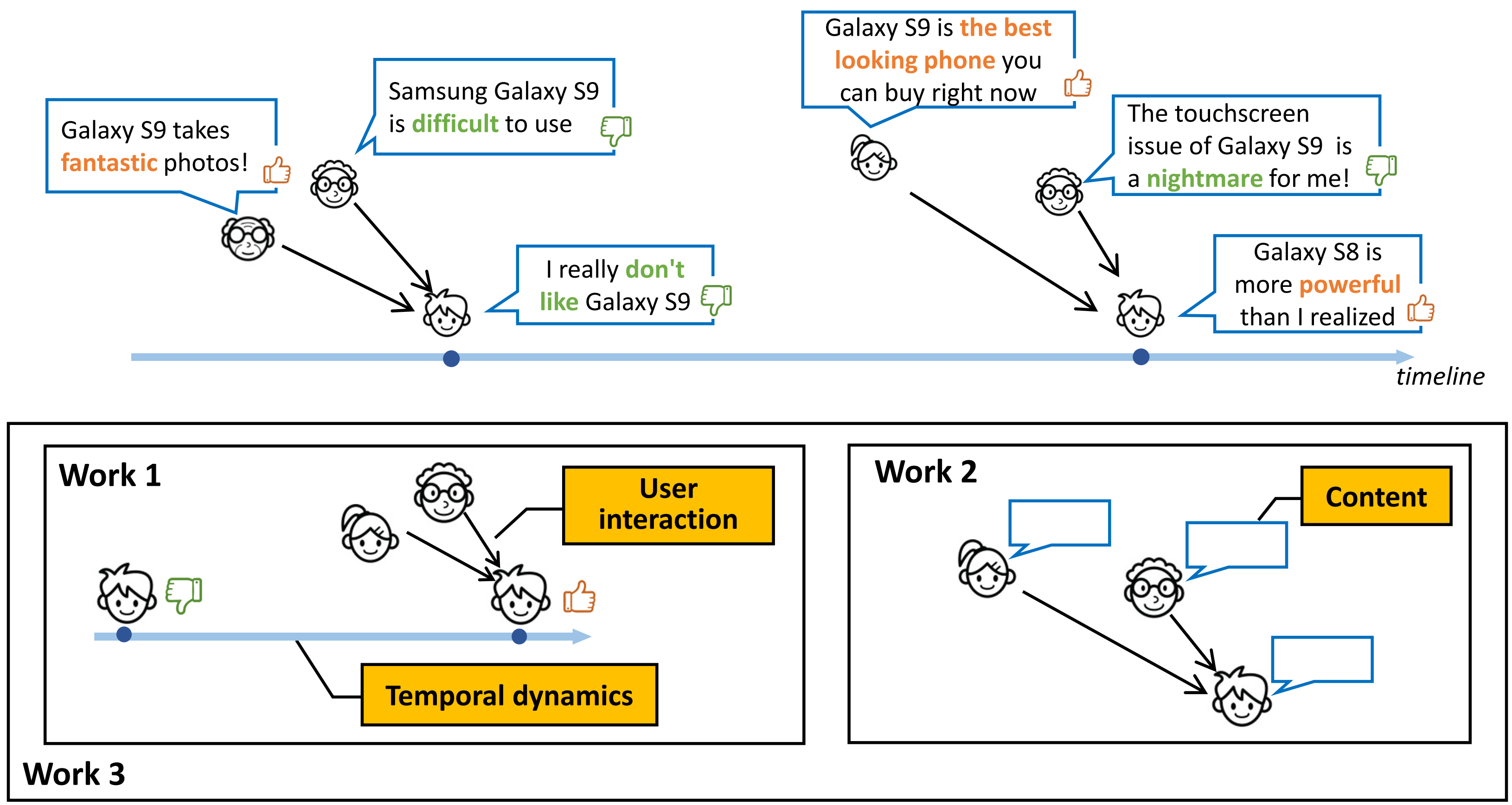


Figure 1.1: The study of opinion influence modeling incorporates three important components: user interaction, temporal dynamic, and opinion content.

estimation can be evaluated using prediction performance.

Work 1: Temporal Opinion Influence Modeling

Previous studies took the opinion influence process as an aggregation process, where a user's updated opinion is assumed to be either the averaged opinions of her/his neighbors, or the majority opinion of her/his neighbors in the previous time step. With these assumptions, the opinion change is simply the aggregation of the opinions an individual receives in the previous time step. Thus, previous studies failed to investigate the temporal dynamics of users' interactions, where a user continuously collects the information from her/his neighbors and repeatedly updates her/his opinions according to the received information. In this work, we propose a temporal opinion influence model [30]. It models opinion dynamics of each user with an individual temporal Markov model and uncovers the opinion influence through correlating the opinion dynamics of connected users. Two indicators are captured to reflect the effect of social interaction on opinion formation, including the friend effect and the opinion effect.

Contributions: We are the first to consider temporal properties of opinion dynamics, and uncover interaction effects in the opinion influence process. The couple Markov chain provides a creative way to uncover the opinion influence through correlating temporal opinion behaviors of users. The experiments on Twitter datasets demonstrate the effectiveness of the proposed model in uncovering opinion influence and in predicting sentiment of future opinion.

Work 2: Content-based Opinion Influence Modeling

On most social media platforms, people exchange information and ideas by posting and replying textual messages. Existing opinion influence studies compress the opinion with a numerical sentiment value, which lacks the exploration of how different expressions affect a user's future opinions. To address the problem, we propose a new opinion influence framework [31]. It is based on the neural network framework and employs the word embedding techniques to deal with the content information [116]. Besides, we further extend the content-based opinion influence study with the focus on the involved users in the opinion influence process. A user may possess two types of identities in a society, including personal identity and social identity. Both identities consist of self-image and affect the opinion influence process. We investigate the opinion influence process by characterizing users with their dual identities. Considering the social identity formation and opinion influence processes both occur in social communication, a novel joint learning algorithm is proposed to simultaneously detect social identities and model opinion influence in a unified framework [32].

Contributions: In this work, content is first introduced in opinion influence modeling. The word embedding techniques provide an effective way to understand exchanged opinions and the proposed NN-based framework well captures the content-based opinion influence process. We also explore the effects of personal images. It is the first time that both personal identity and social identity are considered in the study of opinion influence. The proposed joint learning framework has been demonstrated effectively in capturing the effects of users' dual identities in opinion influence modeling.

Work 3: Content-based Sequential Opinion Influence Modeling

The opinion influence process has been studied by exploring temporal properties or the exchanged content information in previous two works. In this work, we aim to collectively study the above-mentioned components related to opinion influence modeling in a unified framework. It requires the model to temporally track content-based opinion behaviors, and uncover correlations of different users from their dynamics interactions. To this end, we develop a sequential model based on the Recurrent Neural Network framework [33]. Upon this framework, two prediction strategies with different granularities are proposed, including the opinion sentiment prediction strategy and the fine-grained opinion word prediction strategy.

Contributions: This work contributes a novel framework which captures temporal properties of opinion dynamics and the semantic information included in the content. It provides a complete and effective understanding of the opinion influence process, which can be further extended to model the content-based user dynamics in other scenarios. Besides, the fine-grained prediction strategy equips the opinion influence model with the ability to forecast the opinion words other than opinion sentiment, which may benefit more detailed marketing analysis.

1.4 Structure of Thesis

The overall picture of the thesis is as follows. Chapter 1 briefly introduces the background the studies on opinion influence modeling. The research motivation, overview and contribution are also explained. Chapter 2 surveys the existing work on social influence analysis, opinion analysis, and opinion influence modeling. After the brief summary of the literature, we point out the differences between the existing studies and opinion influence modeling explored in the thesis work. Chapter 3 presents the data collection method. Besides, we also design an opinion influence test to verify existence of the opinion influence as the preliminary study. Chapter 4 investigates the opinion influence process from temporal properties of user interactions, where the temporal Markov assumptions are integrated and the effect of social interaction on opinion formation is captured. Chapter 5 explores the content information captured in the opinion influence process. A novel opinion influence framework is proposed to leverage the textual information. Based on the new framework, the dual identities of users are explored and jointly learned in the proposed opinion influence model. Chapter 6 combines the two successful components, including temporal dynamics interactions and content information, in a unified framework. A complete analysis of the characteristics of user behaviors and their opinion influence is conducted. Chapter 7 summarizes the proposed methods, findings, conclusions, and contributions of the work. The potential future extensions of the current work are suggested at last.

Chapter 2

Literature Review

In this chapter, we survey the studies related to this thesis, including social influence analysis, sentiment classification and opinion dynamics modeling.

2.1 Research in Social Influence Analysis

The study of social influence analysis has been particularly active for a number of years in research areas including sociology, physics, marketing and computer science. In these studies, the social media is treated as the platform for increasing the spread of information, products and advertisements over the whole network. The social influence is described as the ability of user to affect the spread of information. Works closely related to social influence analysis on social media involve influencer detection, information diffusion and influence maximization.

2.1.1 Influencer Detection

Finding the influencers over the whole social network has attracted the attention of both academic researchers and business marketers. It focuses on measuring the global influence of users and further finding the top influencers with high global influence. Although there does not exist a unified standard to quantify the user social influence on social media, it is widely agreed that social influence cannot be exerted without social relationships. A series of research studies have examined social influence based on social network. These studies represented a social network as an undirected or directed graph where nodes represented users and edges represented connections between users. Based on the constructed graph, the influence powers of users within the social network were measured through a variety of network properties, such as the in- and out-degree [26], closeness [150] and betweenness [85]. In addition to measuring the authorities of users individually, [95, 81, 99, 24] proposed different approaches to rank user influence by applying graph-based ranking approaches, such as PageRank [129] and HITS [95]. In the meanwhile, considering that social influence varied in users in terms of different topics of interests, some researchers [170, 130] proposed to identify the influential users by taking both graph structure and topical similarity into consideration. It was true that graph structure to some extent reflected social influence of users in the network. However, it was argued that a focus solely on the network established by users cannot depict how influence flows over the network. Thus, the static network structure alone is unable to provide a complete understanding of social influence.

There were some other work measuring and comparing influence of users by considering the topology emerging from the actions of users, including the retweeting and reforwarding actions [125, 128], mentioning actions [26] and clicking actions [42]. Accordingly, the statistics derived from the observed actions, such as the number of retweets [125], the number of mentions [26] were proposed to reflect the influence power of an individual user. Recently, several influence measurement systems including Klout score [139], peerindex [148] have been successfully developed for industry to identify influencers on social media. These systems employed more sophisticated metrics based on the combination of network properties and the actions observed on social media.

Though measurement of users' global influence and detection of top influencers have been extensively studied by researchers and entrepreneurs, and have been successfully applied to empirical marketing activities, the effects of social influence on users' behaviors have not been quantitatively captured in current work. Different from the study on macro-level influence (i.e., global influence), we focus on micro-level influence (i.e., interpersonal influence) in this thesis. Interpersonal influence reflects relationships between each pair of users, and provides opportunities for companies to target their customers with purposes.

2.1.2 Information Diffusion

The diffusion phenomena were firstly emerged in epidemiology. The fast development of social media provided an unprecedented era for new related research directions. Information diffusion on social media is described as a phenomenon that a piece of information spreads along the social network depending on the properties of the edges and nodes. A lot of efforts have been made in order to understand the mechanisms behind the phenomenon. As illustrated in Figure 2.1, each node can be activated by its neighbors with a monotonicity assumption, i.e., the activated nodes cannot be deactivated. Thus, the diffusion process of a specific message can be treated as a sequence of successive activation of nodes through the network [71].

Given the activation sequences, a lot of work have been done to study the mechanisms behind the information diffusion process and further predict the future diffusion path of a specific message. Most of the proposed diffusion models are based on two fundamental diffusion models, i.e., Independent cascade (IC) [61] and Linear Threshold (LT) [68]. The IC model assumes that the newly activated nodes can only activate their neighbors once with the probabilities associated to edges. The LT model defines an influence degree on each edge and the inactive nodes are activated by their activated neighbors if the sum of influence degrees exceeds their own influence threshold. Under the diffusion assumptions of IC and LT models, several

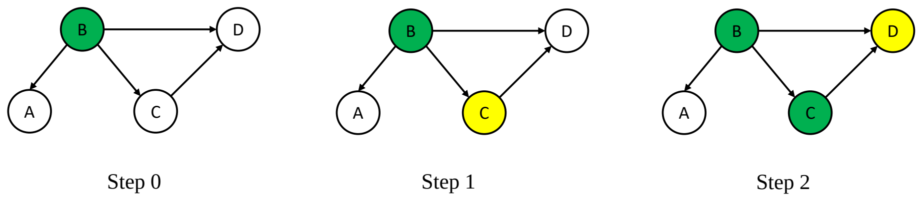


Figure 2.1: Information diffusion process

approaches [62, 123, 143, 144] have formulated the diffusion process with probabilistic generative models. The diffusion influence between users and the diffusion rates were inferred from the observed activation sequences. Though the above methods successfully modeled the information diffusion process, they heavily relied on the diffusion traces to learn diffusion probabilities. Since the number of relationships is much larger than nodes in the network, a large amount of data is needed to learn the diffusion probability between each link. To tackle the sparsity of diffusion paths, Saito et al. [146] proposed to infer diffusion influence based on the attributes of two connected nodes, which were learned from the observed diffusion processes. Following this study, [71] further estimated diffusion probabilities using a variety of features including social, semantic and temporal features.

Previous approaches assumed the existence of network structure, however, in many scenarios, the network where diffusion takes place is in fact difficult to obtain. Rather than predicting diffusion influence between two nodes, Yang and Leveskov [174] focused on modeling global influence a node had on the rate of diffusion and used it to predict the range of future adoption. Recently, motivated by the advent of neural networks, researchers have begun to model the diffusion process from a different viewpoint. A series of RNN-based sequential models [164, 167] were proposed to sequentially encode historical diffusion information as hidden states. The next infected user was then predicted according to the compressed hidden state.

Though information diffusion has been extensively studied by researchers in existing work, it cannot account for users' dynamic behaviors when they continuously change their behaviors under influence. For example, an individual's opinion towards a product would not always be the same. A fan of the product who always holds the positive attitude of the product may become to argue the quality when s/he knows negative news about the product from her/his neighbors. Existing studies on diffusion is information-centric. They use cascade models to learn how information propagates over the network, but less detailed study on the continuous behavioral changes due to social influence. A user-centric influence model is needed to track users' opinion changes and explore opinion influence from dynamic interactions.

2.1.3 Influence Maximization

With the estimates of diffusion influence between users, researchers and marketers have investigated the effects of diffusion influence in viral marketing. Supposing a company would like to promote a product on social media and make a large of amount customers adopt the product, the most cost-efficient way is to select a few number of users to advertise or recommend the product. Under the word-of-mouth effects of diffusion influence, the message of the product can be spread out to trigger a widespread of adoptions. An algorithmic problem, called influence maximization is proposed, i.e., which seed users should be chosen in the initial phase of the word-of-mouth marketing in order to maximize the range of the diffusion?

Domingos and Richardson [49] were the pioneers to solve this problem. Then the problem rapidly became a hot topic in social media analysis. They proposed a probabilistic framework based on Markov random fields. Based on it, several heuristic strategies were suggested to select the users that can influence a large fraction of the network. Kempe et al. [88] first formulated the problem as a discrete stochastic optimization problem and proved that the optimal solution of the problem is at least NP-hard. A naive greedy algorithm was then provided to approximate the problem, which guaranteed the $1-\frac{1}{e}$ approximation ratio. However, the time complexity of the naive greedy algorithm was prohibitive for the large-scale network. There were two reasons for the inefficiency of the algorithm. One was that finding the expected range of the influence spread was a #P-hard problem. The other one was that the greedy algorithm was quadratic in the number of nodes. Considerable work [93, 35, 34, 65, 90, 168] have relied on heuristic techniques to solve the first problem. However, these methods sacrificed the $1-\frac{1}{e}$ approximation ratio of Kempe's approach. On the other hand, a series of approaches [101, 67, 66] focused on solving the second problem to improve the practical efficiency of the greedy algorithm from the implementation view. One notable work was [101], which exploited submodularity to develop an efficient algorithm called CELF, based on a lazy-forward optimization. Recently, [21] made a theoretical breakthrough by improving the quadratic nature of the greedy algorithm to decrease the upper bound of running time. Several follow-up work were developed to improve the practical efficiency while retaining the same worst-case guarantees of running time [156, 165, 72].

Influence maximization is an optimization problem, which aims at maximizing the effects of influence in marketing activities. Instead of capturing influence from the observed behavioral data, they take interpersonal influence in terms of diffusion probability as the prior knowledge. The purpose of this thesis is to measure interpersonal opinion influence from dynamic user interactions, which can be regarded as prerequisite of influence maximization studies.

Summary: Although social influence has been extensively studied by researchers from a variety of viewpoints, how opinion influence affects users' dynamic behaviors during interactions is still an open question. In this thesis, we study social influence from user dynamic interaction point of view and explore its effects on forecasting users' future behaviors.

2.2 Research in Sentiment Classification

With the advent of Web 2.0, people have opportunities to share their points of view regarding daily activities, hot issues and etc. on different types of platforms. For instance, Epinion[1] allows customers to write pros and cons regarding their personal experiences when using products and IMDb[2] enables people to write reviews about the movies they watch. The published opinions benefit a variety of applications including public mood observation, politics election forecast, and marketing strategies design. To accomplish these tasks, researchers and enterprisers rely on sentiment classification, which aims to determine the polarity of sentiment (i.e., negative sentiment, positive sentiment and neutral sentiment) from textual messages. Basically, the studies of sentiment classification proceeded along two tracks as reported in [114]. One is machine learning based approaches and the other is sentiment lexicon based approaches.

The machine learning based approaches formulate sentiment classification as a text categorization problem [2], where the types of sentiment polarities are taken as the classification labels. In these approaches, the document is represented by a variety of features, and machine learning algorithms are applied to determine the sentiment polarities of documents. Pang and Lee [133] were the pioneers to employ three classification methods, i.e., Naive Bayesion (NB) [57], Maximum Entropy (ME) [16] and Support Vector Machine (SVM) [151] to determine the positive and negative sentiments of movie reviews. The classifiers were trained using N-gram features.

Later, more sophisticated learning methods and richer features were devised to improve the performances of classification. As for features, researchers have explored sentiment words [91], Part-Of-Speech (POS) tags [122, 171], negation words

[1] www.epinion.com

[2] www.imdb.com

[124, 89], contextual valences and sentiment shifters [89, 106]. Other than directly applying basic machine learning algorithms, researchers also developed specific methods for sentiment classification. McDonald et al. [113] presented a structured model to jointly classify the sentiments at both sentence and document levels. With the idea of determining sentiment from fine-to-coarse, a hidden conditional random field approach [138] was developed. [18] developed a decision tree of SVMs for document-level multi-class sentiment classification, which leveraged the inter-class similarity in the learning process. Since the performance of a machine learning algorithm heavily depended on the choice of data representation, a novel type of document representation arose with the advent of deep learning. The documents were condensed into low-dimensional feature vectors. Several deep Neural Network (NN) models have been successfully applied to identify sentiment polarities from dense representations [92, 155, 153, 141, 70, 69]. The NN-based approaches gained significant improvements compared with the traditional machine learning methods.

Different from machine learning based approaches, which learn sentiment from the large scale of labeled text data, lexicon-based approaches rely on sentiment lexicons to determine the sentiment polarity of a document/message. It aggregates the sentiment orientations/strengths of words or phrases in the document/message [162] to derive the overall sentiment of the whole document/message.

In lexicon-based approaches, the sentiment lexicon is crucial to the quality of sentiment classification. A lexicon is a list of opinion words associated with their sentiment polarities or strengths. It can be constructed manually or using seed words to expand. Given the lexicon, the sentiment of a document can be obtained by aggregating the sentiment intensities of all opinion words occurring in a document [137, 48, 152]. LIWC sentiment lexicon [3] and Harvard GI lexicon [4] are two popu-

[3] www.liwc.net

[4] http://www.wjh.harvard.edu/ inquirer/

lar well-established sentiment lexicons constructed by human for sentiment analysis. They have been widely used in a variety of applications, including differentiating couples in happiness from couples in sadness based on their textual instant communication [73], predicting depression states of users in social media from their posted messages [45] and etc. Differently, Hu and Liu [79] annotated words with positive or negative sentiment orientations with a bootstrapping process. Baccianella et al. [8] constructed a lexicon where words were annotated with numerical scores relating to positivity, negativity, and objectivity using a mix of semi-supervised algorithms. Despite the widespread of these lexicons in detecting sentiment for different applications, they did not consider the microblog-like textual features, such as emoticons (e.g., ":-)" denotes a "smiley face" indicating a positive sentiment), initialisms (e.g., "WTF") and slang (e.g., "nah") [83]. Recently, Hutto and Gilbertcite [82] constructed a lexicon specifically tuned for microblog-like contexts using a combination of qualitative and quantitative methods. It included Western-style emoticons, commonly used initialisms and slang, and was verified manually. The experiments on the Twitter data set prove that sentiment classification achieved a higher accuracy based on this lexicon compared with other existing lexicons.

Machine learning based approaches require a large scale of annotated data to train classifiers. The learning process is time-consuming. Unsupervised lexicon-based approaches are time-saving and easy to be adapted to the datasets in different domains when an appropriate lexicon is selected [152]. Thus, we decide to use the lexicon-based approach to determine opinion sentiments included in social media messages.

Summary: Sentiment analysis is a popular research area in recent two decades. Its focus is to identify the sentiment polarity of a document or message. However, it is not able to discover the inherit mechanism of how the sentiment of opinion forms or changes during user interactions. In this thesis, we take the sentiment as the premise and exploit the observed opinion behaviors to study the opinion influence process which shapes people's opinions.

2.3 Research in Opinion Dynamics Modeling

In a social environment, an individual would not always hold the same opinion on a specific issue. Her/his opinion will change after receiving opinions from her/his neighbors. The change of opinions towards of an issue is called opinion dynamics. The research on opinion dynamics has been studied for decades of years in the areas of physics, psychology and sociology. In the earlier studies, the term "opinion" was referred to as the belief or judgment of a person towards an issue. It can be either the emotion expressed in a picture or the sentiment polarity towards a product. Researchers started with the laboratory experiments to testify if a person actually changed her/his answer after looking at the answers from the other people. A great number of extensive theoretical research works were developed to characterize opinion dynamics of agents in a social system with different influence assumptions. In what follows, we introduce these laboratory experiments and review the existing explorations on influence assumptions for opinion change.

2.3.1 Laboratory Study

Asch [7] was the first to conduct experiments and proved the effects of social influence in shaping people's opinions. In his seminal experiments, Asch observed that a large fraction of people tended to accept the opinion that everyone else in the group said the same, even if their answers in the beginning were self-evident. He claimed that the tendency to conform to others' opinions in the society was very strong. In Asch's experiments, all subjects could know the opinions of all the other people. Deutsch and Gerard [47] argued that the observations may result from the two aspects of conformity. One was normative influence, where the subjects conformed to gain the acceptance of the group, and the other was informational influence, where subjects agreed the responses of others and updated their own opinions. To demonstrate informational influence, they developed similar experiments where the other answers shown to the subject were anonymous and fabricated, and observed similar behavior as Asch. It indicated that even only under information influence and no peer pressure, people were still influenced by others to change their beliefs.

Though these experiments provided evidence that people actually changed their beliefs after receiving the opinions from their neighbors, they did not provide a quantitative model to uncover what opinion of a subject will change to during the communication.

2.3.2 Opinion Dynamics Model

To further study the mechanism of how people update their opinions within a social system, a variety of opinion dynamics models were designed. Most existing work in this line of research agreed that the opinion influence process was an aggregation process. Relying on the pre-defined ending opinion status of the system, e.g., opinion consensus and opinion polarization, different opinion updating rules were proposed. Basically, two types of opinion updating rules were popularly applied to model the opinion influence process, one was the averaging rule [46, 75] and the other was the voter rule [40, 77, 97].

Averaging Model: The Degroot model was the first proposed averaging model [74, 46]. It modeled the opinions of individuals as continuous variables and assumed that a person formed her/his new opinions according to a *rule of thumbs* method. It assumed that an individual took the averaged opinions of neighbors to update her/his own's [159]. Under the averaging rule, the opinions of all users in the social system would reach a consensus at last. Starting from the Degroot model, there was a developing line of studies adjusting the averaging rules in accordance to different assumptions towards the social system. Friedkin and Johnsen [56] argued that consensus was rare in society. Usually, public opinions tended to be in a state of persistent disagreement. They extended the Degroot model to include both disagreement and consensus through assigning an intrinsic opinion for each user in addition to the opinions they expressed. Furthermore, the Flocking model [75], which considered the conformity bias [109] was proposed. It was natural that an individual paid more attention to those neighbors who had more similar beliefs with her/himself when adopting their opinions. The Flocking model first selected an individual's trusted neighbor set and then the same updating rules used in the Degroot model were employed. [111] demonstrated that opinions of the social system also reached a consensus under the Flocking model. Following the averaging updating rules, a large amount of studies have been done by considering misinformation [1], stubborn agents [60] and etc.

Voter Model: Different from the opinion dynamics model with the averaging influence rule, Clifford and Sudbury [40], and Holly and Liggett [77] proposed the voter model independently. They represented opinions with discrete values, usually binary opinions like supporting or opposing. The assumption of the voter model was that a person selected only one of her/his neighbors uniformly at random at each timestep, and took the neighbor's opinion as her/his new opinion. A modification of the voter model with the majority rule was proposed by [97], where a person adopted the major opinion of her/his neighborhood. Under the voting rules, the model has been proved to reach to a consensus [1]. Several variants of the voter model were proposed later. Yildiz et al. [179] proposed a modification to consider stubborn agents. Kempe et al. [87] proposed a modification where both passive influence and initiative selection were integrated in a unified framework.

Hybrid Model: Recently, Das et al. [43] conducted three online user experiments to distinguish the Degroot model, Flocking model and Voter model. They observed that none of the three models could comprehensively uncover people's opinion behaviors. Thus, they proposed a hybrid model named biasedvoter model to take the benefits of three models. They assumed that the opinion of each person was driven by the following three forces: stubbornness, the Degroot behavior that an individual averaged the opinions of her/his neighbors, and the biased conformity that a user tended to randomly select the opinion of a person who had the similar opinion with her/himself.

So far, we have examined three types of opinion dynamics models with different opinion updating assumptions. Though they provided meaningful understanding on opinion dynamics, all of the theoretical methods designed the updating rules under the assumption of the ending opinion status (i.e., consensus or divergence). However, during the interactions in reality, users continuously change their opinions according to the received information. It is difficult for a group of users on the social media to achieve a steady state. Besides, as the behavioral data is difficult to obtain, existing studies only test their models on the simulated dataset. They neither fit the data collected from a real scenario where people change their opinions towards an issue with different opinion dynamics model nor provided quantitative results to demonstrate the predictability of different approaches with different influence rules. De et al. [44] first made no assumptions on convergence of opinion. They studied the transient opinion dynamics and proposed the asynchronous linear model (AsLM) with unbounded interpersonal influence to model opinion behaviors on social media. They also verified the effectiveness of the model on the real social media dataset.

Summary: Though De et al. [44] extended traditional theoretical study of opinion dynamics to practical validation on real-life social media data, they still simplified user interactions and ignored many valuable features affecting opinion formation. Opinion influence arises when people dynamically interact with each other. The temporal properties of user interactions have not been studied in the existing work. In the meanwhile, people exchange opinions through the User-Generated Text (UGT), and the effects of the content information on opinion formation has not been studied in the previous work. In this thesis, we explore different characteristics of social communication and conduct empirical studies on opinion influence modeling.

Chapter 3

Data Collection and Influence Test

The main purpose of this thesis is to uncover opinion influence arisen from user communication traces on social media. To understand opinion influence in an empirical setting, a dataset which contains the opinion records and the users' network structure is necessary. In the psychology area, researchers designed a variety of laboratory experiments to collect the shift of opinions when a person is exposed to the responses of other participants. After analyzing the data collected in the laboratory experiments, researchers demonstrated that opinion influence actually exists in social communication, and changes future behaviors of people [7, 43]. However, the designed in-house experiments have two significant drawbacks. First, the carefully designed experimental setting simplified the communication in the real world. Moreover, the sizes of the samples used in the laboratory experiments were very small, which made it difficult to verify opinion influence. Thus, it is necessary to design a large scale of experiments to demonstrate the existence of opinion influence. As social media provides a convenient way for researchers to observe and to collect communication traces of users, we choose the popular social media platform Twitter to study opinion influence in this thesis. In this chapter, we first introduce the method we use to collect the opinion records of social users from Twitter, and then report a statistical test result that verifies the existence of opinion influence.

3.1 Data Collection

Understanding opinion influence is of great importance for business promotion. We focus on the study of discussions towards business products, which provides insights into a future commercial strategy. As one of the most popular social media platforms, Twitter allows users to publish their own content or opinions about a specific product, as well as establish relationships with other Twitter users. It is a great platform for us to study user communication and opinion influence. We choose three well-known electronic products widely discussed on Twitter as the example topics to study opinion influence. They are Samsung Galaxy, Xbox and PlayStation. Twitter provides a variety of APIs for researchers and developers to collect data. Considering the collection of the historical data is constrained by Twitter [1], we use the streaming API [2] to collect the data in a tracking fashion. We spend totally 8 months to collect the related tweets published from 31st March, 2014 to 30th November, 2014.

The dataset is built for each product separately in the following way. All English tweets containing the topical word like Xbox that occurred during the above-mentioned time period are collected. Each tweet is associated with the id of the user who posts it, the posting time and the textual content. We carefully preprocess the collected tweets to construct the final collection of tweets as below:

- Discard the retweeted tweets
- Remove the non-English tokens included in the text
- Discard the tweets containing less than 5 tokens

Based on the preprocessed tweet collection, we build the user set, which contains all the users who are interested in the specific product. Because opinion influence

[1] https://developer.twitter.com/

[2] https://dev.twitter.com/streaming/overview

Table 3.1: Network statistics.

Topic	Samsung Galaxy	Xbox	PlayStation
# of users	8921	4358	5158
# of avg friends	14.42	9.58	11.83
# of avg tweets	87.31	62.92	150.00
activity level	0.39	0.37	0.28

is exposed through the long-term communication, a certain number of tweets from users are necessary for us to capture influence. Among all users who published the related tweets during the period, we construct the active user set for each product by filtering out the inactive users who posted less than 30 tweets and the over-active users who published more than 1000 tweets. In addition to the published tweets, the user network is another important data source to study opinion influence. The user relationship is obtained via the "get friends" function [3]. Finally, we obtain both users' posting records and their connected network.

Table 3.1 summarizes the statistics of the datasets. "# of users" describes the size of the network, and "# of avg friends" represents the average number of Twitter users an individual user follows. Because different products may have different degrees of attractiveness to consumers, we analyze the activity level of each product. The communication process can be divided into a number of communication rounds. Each communication round starts after the user posts a tweet, and ends when the user updates her/his opinion and posts a new tweet. During one communication round, not all of a user's friends provide the suggestions. We define the friends who actually post tweets and may influence the user's next opinion as the active friends. Further, for each communication round, we calculate the percentage of the active friends and average the value of all users. The result is the "activity level", which implies the user involvement in social communication on each product.

[3]https://developer.twitter.com/en/docs/accounts-and-users/follow-search-get-users/api-reference/get-friends-list

3.1.1 Opinion Analysis

To study opinion influence, a comprehensive analysis on posting tweets is needed. In this section, we first introduce the method employed to identify the positive, negative and neutral sentiments of opinion. Then we explain how to extract the opinion words from the tweets. Note that, the terms sentiment and sentiment polarity are used interchangeably in this thesis.

Sentiment Classification

Sentiment identification has been widely studied by researchers and engineers. Most of existing ready-made sentiment classification tools are tailored for the long texts, which are not suitable for detecting the sentiment of short texts. Here, we employ a sentiment analysis method [82] proposed by Hutto and Gilbertto, which was carefully devised to detect the sentiments of tweets, and has been proved to achieve a high accuracy on detecting the sentiment of tweets. In this approach, a sentiment lexicon was specifically designed to fit for the Twitter-like context. It associated each opinion word with its sentiment intensity. The lexicon was validated by human annotators. Based on the lexicon, they proposed a method called Vader to obtain sentiment score of a tweet by combining sentiment intensities of all appearing opinion words of the tweet. According to the obtained sentiment score s of each tweet, its sentiment polarity o could be determined as follows.

$$o = \begin{cases} +1 & \text{if } s > \epsilon \\ -1 & \text{if } s < -\epsilon \\ 0 & \text{otherwise} \end{cases}$$

According to the experimental results listed in [82], the accuracy on 4000 randomly selected tweets is 96%. It also outperforms the methods based on typical sentiment

Table 3.2: Sentiment statistics.

Topic	Samsung Galaxy	Xbox	PlayStation
% of negative sentiment	11.40	16.33	12.76
% of positive sentiment	19.96	40.09	24.46
% of neutral sentiment	68.64	43.58	62.78

lexicons, such as LIWC [136], SentiWordNet[4], SenticNet[5].

We follow Hutto and Gilbertto to detect the sentiment polarities of tweets in the collected dataset. The percentage of each type of sentiment polarity is presented in Table 3.2.

Opinion Word Extraction

The sentiment polarity summarizes the opinion of a tweet. To understand opinions in more detail, the content in tweet is important. For each tweet, we extract its opinions words according to the Twitter-specific sentiment lexicon used in Vader [82] with the following rules. We first define a negation list which contains negation words and phrases, such as “don’t”, “didn’t”, “never so” and etc. When we detect an opinion word, we find whether it follows a word or phrase included in the negation list. if yes, we retain the phrase “not”+“opinion word” as the opinion word of the message. For example, the opinion word extracted from the tweet “I don’t like the Samsung Galaxy S6.” is the phrase “not like”. For the tweets only stating the facts without expressing an opinion, we use the word symbol “NeuW” to represent them. To alleviate the word sparsity problem, we only keep the opinion words that occur more than 50 times in the whole dataset and replace the infrequent opinion words with the corresponding symbols. The positive opinion words are replaced with the symbol “PosW”, and the negative opinion words are replaced with the symbol “NegW”.

[4]http://sentiwordnet.isti.cnr.it/

[5]http://sentic.net/

Finally, the numbers of the remaining opinion words for the topic "Samsung Galaxy", "Xbox", and "PlayStation" are 880, 1146 and 505, respectively. The percentage of tweets containing the opinion words for the topic "Samsung Galaxy", "Xbox", and "PlayStation" are 57.7%, 68.4%,62.8% accordingly.

3.1.2 Variation across Datasets

Though the three topics we select belong to the same product category, they have different characteristics in terms of the network structure and communication. From the network statistics and sentiment statistics listed in Table 3.1 and 3.2, we find that the involvement of users and their sentiment diversity are different for the three products. "Samsung Galaxy" has the largest customer network which shows that the product is very popular on Twitter. It also has a high activity level, which means that users are actively involved in the discussion of the product. Different from "Samsung Galaxy", "Xbox" has a relatively small group of potential customers, but their communication on the product is very frequent. It may indicate that the users interested in "Xbox" form a small but tight network. Though "PlayStation" attracts a certain number of users, their interests on this product is less compared to the other two products. Thus, in the following sections, the experiments on the three products demonstrate the robustness of the proposed models on products with different characteristics.

3.2 Opinion Influence Test

In Chapter 2, we review several studies on the verification of opinion influence in the laboratory test. However, on social media, where users obtain information from a variety of information channels, the existence of social opinion influence has not been demonstrated through a statistical test. A popular shuffle test has been proposed to testify the existence of diffusion influence during information propagation [4]. The basic assumption is that if diffusion influence does not exist in the social network, the time a user reposts a message should be independent of the time when her/his friends repost the same message. Though the shuffle test has been successfully applied to demonstrate that a specific message actually diffuses over the network through the connected relationship, it cannot tell whether a user changes her/his mind according to her/his neighbors. Inspired by the idea of the shuffle test, we propose an approach to demonstrate the existence of opinion influence on social media.

Before introducing the details of the proposed opinion influence test, we need to clarify the definition of opinion influence first. In social communication, opinion influence can be observed when an individual expresses a new opinion, which is different from her/his prior opinion. We devise the opinion influence test under the following assumption. We assume that if opinion influence does not exist in the social network, one's opinions should be independent of her/his friends' opinions. In other words, the correlations between one's opinions and her/his friends' opinions should keep the similar properties after the users' opinion sequences are shuffled. Under this assumption, we present the details of the opinion influence test below.

First, the instances of opinion changes are extracted from the opinion sequences. When a user u posts a tweet with the sentiment different from her/his prior sentiment, the tweet is taken as one instance of opinion change. Among all the instances of opinion change, we find those opinions that may be changed by users' neighbors and call them the activation instances. The activation instances are defined as follows: if we observe that u's neighbor posts an opinion o different from u's current sentiment, and then u changes her/his mind and posts a new message with the sentiment o, we say that u is activated by her/his friends, and the instance of opinion change is taken as an activation instance. In the whole dataset of a topic, the percentage of the activation instances among the instances of the opinion change reflects the correlation value between users' opinions and their neighbors' opinions. Besides the

Table 3.3: The correlation between a user's opinion and neighbors' opinions

Topic	Samsung Galaxy	Xbox	PlayStation
shuffled dataset	0.2640	0.3457	0.2649
original set	0.3152⋆	0.3958⋆	0.3296⋆

⋆ represents a significant higher value

original dataset, 100 shuffled datasets are constructed for the statistical test. The shuffled dataset is constructed by randomly permuting the connections between the times and tweets included in the posting histories for each user. The correlation value is calculated on each shuffled dataset, and we perform the t-test to verify whether the population mean of all shuffled datasets is significantly different from the correlation value obtained from the original dataset.

The results are listed in Table 3.3. We can observe that the correlation value on the original dataset is much larger than the population average of the shuffled datasets, and the increasement is significant. There indeed exists a correlation between a user's opinions and her/his neighbors' opinions in the original dataset compared with the shuffled dataset, which demonstrates the existence of opinion influence on social media.

Chapter 4

Temporal Opinion Influence Modeling

4.1 Chapter Overview

The phenomenon that an individual could be influenced by neighbors to update her/his previous opinion has been first observed in the psychological experiments [47, 7]. After that, an extensive range of theoretical studies tried to model the opinion influence process under different assumptions. Unfortunately, due to the difficulty of collecting data from empirical communication records, their studies were limited to the computer simulation only, and did not actually capture opinion influence in an empirical setting. The advent of social media provides a channel for researchers to observe every change of people's opinions during social communication. In this work, we aim to uncover the opinion influence process by empirically tracking people's opinion dynamics.

Most existing studies simplified the opinion influence process as an aggregation process. An individual was assumed to either take the averaged value of her/his neighbors' [46, 75], or randomly select an opinion from neighbors' previous opinions [40, 77] as the next updated opinion. The aggregation process considers the exchanged opinions in a short time period only, and does not study the emergence of opinion influence during temporal interactions. To better understand how opinion influence arises during the communication, we describe the emergence process of opinion influence as follows. In the beginning, users who do not acquaint with each other and little influence can be observed. After interacting with each other, users gradually know each other. Then influence comes out and can be observed through the temporal changes of their opinions. Thus, we aim to estimate opinion influence by tracking the temporal dynamics of users' opinions driven by their social interactions.

To this end, we develop a novel temporal opinion influence model (TIM) inspired by the success of coupled HMM in modeling interacting components. TIM has the ability to temporally track the dynamics of every individual's opinion behaviors, and model the interactions between connected users. The opinion behaviors of each user are individually modeled as a Markov chain and each state of the chain represents the sentiment of opinion a user posts at each timestamp. Two temporal Markov assumptions are proposed to capture the individual temporal properties and to correlate individual's opinion chains with the opinion chains of connected users. To better describe the interactions between users, we propose two indicators to capture the effects of social interactions. They are the opinion effect and friend effect. The proposed model achieves better performances compared with existing opinion influence models on the Twitter dataset. It demonstrates that understanding temporal interactions is very important for uncovering the opinion influence process, and it is well captured by the proposed model.

4.2 Related Work in Coupled HMM

In the camp of computer vision, a series of coupled Hidden Markov Model (HMM) has been successfully applied to model the interacting components within a person/system. The coupled HMM takes the advantages of HMM on tracking the dynamics of every single component, and it has the ability to capture the interactions between different components. It was first proposed by Brand et al. [23] to recognize the T'ai Chi gestures, which involved interactions between two hands. They modeled the activities of each hand as a Markov chain and tried to understand what a person is doing based on the interactions between two hands. Later, several extensive models based on the coupled HMM were developed to recognize more complex actions that performed with multiple components. Ren et al. [140] proposed a primitive-based coupled HMM, which attempted to understand a teacher's intention from her/his two upper-limbs interactions. A coupled hidden semi-Markov Model was used to predict what event would happen by modeling trajectories of moving people, and the results could be integrated into the outdoor visual surveillance system [126]. Though these studies successfully proved the ability of coupled HMMs in modeling the interacting components within a system, their objective was to understand the state of the whole system through the interactions. The dynamic properties of every single component were not sufficiently studied.

Following this idea, a simplified coupled-HMM model, Influence model [6] was theoretically studied by Asavathiratham, and was employed for understanding the behaviors of a large number of interacting components in a complex network, such as communication network, transportation systems, and power girds. Basu et al [11] employed the concept of the influence model to learn the speaking/silent action behaviors of participants in a small scale of laboratory experiments. Pan et al. extended the influence model to infer how influence changes dynamically [132]. However, most of existing studies assume that each participant is aware of the state changes of others, and the next actions/states of all participants can be updated synchronously. The assumption seems reasonable when the binary actions (e.g., silent or speaking) could be observed at any time. However, modeling online social behaviors,

Table 4.1: Definition of notations

Notation	Description
V	a set of $\|V\| = N$ users.
F_u	a set of u's neighbors in the network.
$o_u(i) \in O$	the sentiment of opinion u posts at timestamp $t_u(i)$.
$O = \{1, 2, 3\}$	represents negative, neutral and positive sentiment accordingly.
$S_u(< i) \triangleq \{o_u(1), o_u(2), ..., o_u(i-1)\}$	u's opinion sequences before the timestamp $t_u(i)$. The opinions are sorted according to posting time decreasing, i.e., it satisfies $t_u(i) > t_u(i-1)$.

like opinion behaviors, is different. Users have different habits of using the social media, and it is difficult to divide their posting records into the sequences with the fixed time intervals. Thus, the specific properties of the opinion behaviors should be considered in opinion influence modeling.

4.3 Problem Formulation

Formally, we represent the network of users who are interested in a specific product as a directed graph $G = (V, E)$, where each vertex $u \in V$ represents a user, and each edge $(u, v) \in E$ represents a neighbor relationship meaning that u follows v. The total number of users is N. Derived from G, a neighbor set is constructed for each user $u \in V$, which is denoted as $F_u = \{v|(u, v) \in E\}$. The size of F_u is m_u.

Besides the network structure of users, we also define the notations of the posting records. For each user u, we formulate her/his opinion posted at timestamp $t_u(i)$ as $o_u(i) \in O$. In this thesis, $O = \{1, 2, 3\}$ i.e., each opinion $o_u(i)$ can be 1, 2 or 3 indicating the negative, neutral or positive sentiment respectively. The posting history of each user u can be represented as a sequence $S_u = \{o_u(1), \cdots, o_u(n_u)\}$

with the size n_u.

Given the set of neighbors set F_u and the opinion records $S_u(< i)$ for every user u, our objective is to predict u's next sentiment of opinion $o_u(i)$ at time stamp $t_u(i)$.

4.4 Proposed Model

A temporal opinion influence model is described in this section. It has the ability to track the opinion dynamics of users and temporally capture the correlation between users. Considering the specific characteristics of social interaction, two assumptions are integrated into opinion influence modeling, including the temporal Markov assumption and the interaction-based opinion influence assumption. Given the proposed model, we can further predict sentiment of the future opinion for each user.

For each user u, we use an individual Markov chain to model u's opinion behavior, and each state in the Markov chain represents the sentiment a user posts at each time stamp. The evolution of the Markov chain is affected by the states of its neighboring chains. To capture the opinion influence from a user's neighboring chains, the traditional state transition based on a single chain is transformed to a joint transition probability. Because users post the messages in different times, the opinion state of each Markov chain is updated individually. The opinion influence process is formulated by a joint state transition in 4.1, which describes the probability that u posts the opinion $o_u(i)$ at the timestamp $t_u(i)$ under the influence of her/his previous opinions and her/his neighbors' opinion before $t_u(i)$.

$$P(o_u(i)|S_u(< i), S_1(< i), \cdots, S_v(< i), \cdots) \tag{4.1}$$

where $S_v(< i)$ represents the opinions that u's neighbor $v \in F_u$ posts before $t_u(i)$.

To further model temporal properties of users' opinion changes and capture dynamic interactions, we put forward the following additional assumptions. They are

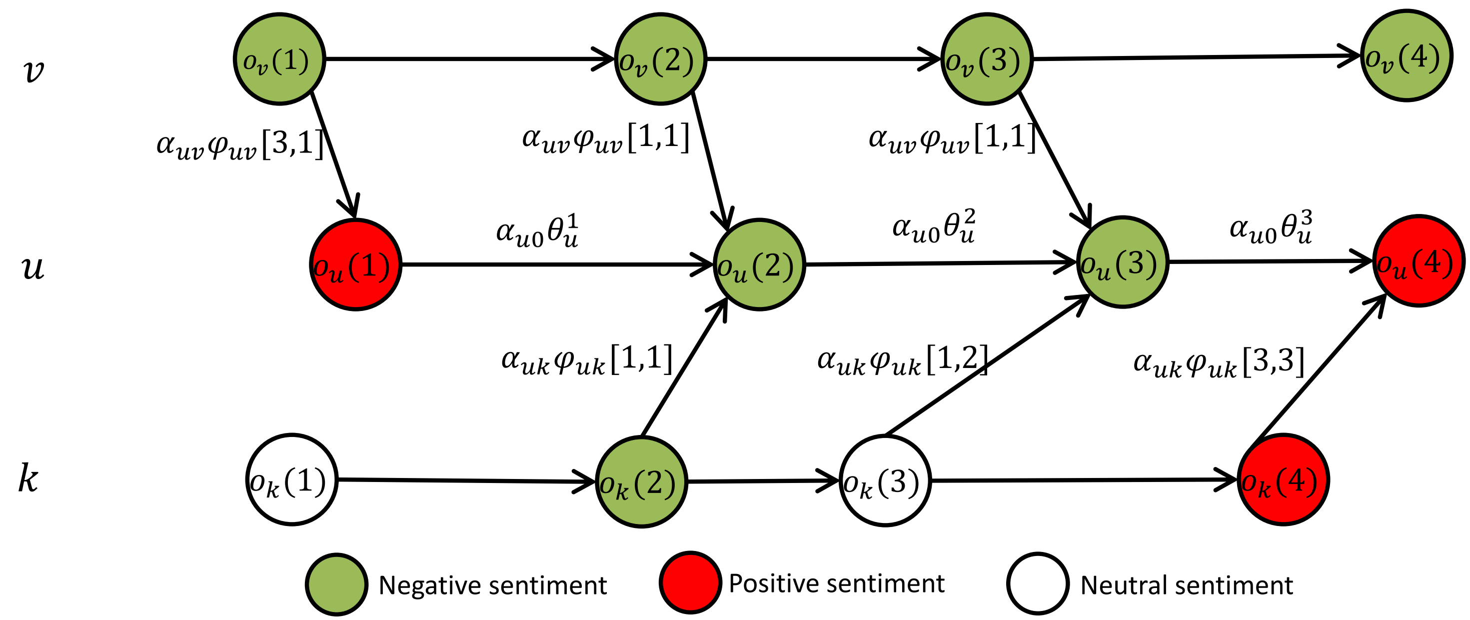


Figure 4.1: The graphical representation of temporal influence model where u follows v and k. It presents how u changes its opinions under the influence of v and k

temporal Markov assumption and interaction-based opinion influence assumption. The graphical illustration of the proposed model is presented in Figure 4.1.

Temporal Markov Assumption

When a user posts tweets about a specific topic, s/he will obviously express opinions following her/his personal prior thought, which means a user prefers to insist on her/his tastes when writing a new one. In this situation, the traditional 1-st order Markov assumption is no longer applicable. In TIM, we assume that one's future sentiment is relevant to the previous personal preference. Hereby, the state transition probability of the individual Markov chain can be represented as a moving vector $\boldsymbol{\theta_u^i}$, which is updated when a user posts a new opinion.

$$p(o|S_u(< i)) = \boldsymbol{\theta_u^i}[o] = \frac{\sum_{l=0}^{i-1} [o_u(l) = o]}{i-1} \tag{4.2}$$

where the symbol $[x]$ equals to 1 when x is true, and 0 if it is false. $\boldsymbol{\theta_u^i} \in \boldsymbol{\Theta}$ satisfies: $\sum_o \boldsymbol{\theta_u^i}[o] = 1$, where $o \in O$.

In addition to following personal preference, a user also takes the opinions from her/his neighbors as the reference to change her/his future sentiment. Usually, it is the most recent message from one's neighbor that triggers the user's change. The joint state transition probability can then be simplified with the temporal assumption, i.e.,

$$P(o_u(i)|S_u(< i), S_1(< i), \cdots, S_v(< i), \cdots) = P(o_u(i)|S_u(< i), o_1(t_1), \cdots, o_v(t_v), \cdots) \tag{4.3}$$

where $o_v(t_v)$ represents the opinion that u' neighbor $v \in F_u$ posts after u's previous opinion and before u's current opinion, which satisfies $t_u(i-1) < t_v < t_u(i)$.

Interaction-based Opinion Influence Assumption

In addition to the temporal assumptions, we also propose two indicators to capture the effects of social interactions on a user's opinion formation. They are friend effect and opinion effect.

When a user adopts the information received in the social network, s/he will inevitably consider the sources of the information and be influenced by some neighbors more than others. The friend effect, which denotes interpersonal influence between users has been widely employed in existing opinion influence models to interpret the correlation between user pairs. We use the parameter $\boldsymbol{A}$ to interpret the friend effects.

$$\boldsymbol{A} = \{\alpha_{uv} | (u \in V, v \in F_u | v = 0)\} \tag{4.4}$$

where α_{uv} represents the strength of influence v exerts on u. When $u = v$, we use α_{u0} to represent personal insistence on one's preferred opinion. The influence strength for each user satisfies the condition of $\sum_v \alpha_{uv} + \alpha_{u0} = 1$.

Though the friend effect is able to capture the ability a user possesses to influence another, it cannot completely reflect opinion influence during interaction. The opinion effect, which captures the effects of different opinions, is also important. For example, if user u takes her/his neighbor v as an enemy, which means everytime when v praises a product, u may oppose v and express a negative opinion towards the product, and vice versa. In this situation, the influence of v is strong, but the effect of the opinion is negative. In the framework of Markov chain, the opinion effect could be modeled through the state transition matrix, which represents the probability that a Markov chain changes from one state to another. Because the single Markov chain could only capture the transition within the independent opinion sequence, it lacks the ability to uncover the influence effects among different user dynamics. To uncover cross-sequence dependencies and capture effects of opinions between different users, we propose to correlate the Markov chains of all users through their network structure, and construct a new type of state transition matrix $\mathbf{\Phi}$ to account for opinion dynamics.

$$\mathbf{\Phi} = \{\boldsymbol{\varphi_{uv}} | u \in V, v \in F_u\} \tag{4.5}$$

where $\boldsymbol{\varphi_{uv}}$ is an $|O| \times |O|$ matrix describing the different effects of v's sentiment on u's future sentiment.

$$\boldsymbol{\varphi_{uv}} = \begin{array}{c|ccc|} u \backslash v & 1 & \cdots, & |O| \\ \hline 1 & \boldsymbol{\varphi_{uv}}[1,1] & \cdots & \boldsymbol{\varphi_{uv}}[1,|O|] \\ \cdots & \cdots & \cdots & \cdots \\ |O| & \boldsymbol{\varphi_{uv}}[|O|,1] & \cdots & \boldsymbol{\varphi_{uv}}[|O|,|O|] \end{array}$$

where $\boldsymbol{\varphi_{uv}}[o, o']$ represents the probability that u expresses the sentiment o conditioned on v's sentiment o', and the elements in each column satisfy $\sum_o \boldsymbol{\varphi_{uv}}[o, o'] = 1$, for $o \in O$.

Given the two indicators of opinion influence, i.e., friend effect and opinion effect, we convert the joint state transition probability in Equation 4.3 by following the linear property of opinion influence [44]. It means that each user is independently influenced by her/his neighbors, and the joint state transition probability could be calculated as the linear combination of the individual opinion transition probability between each pair of users.

$$P(o_u(i)|S_u(< i), o_1(t_1), \cdots, o_v(t_v), \cdots) = \sum_{v \in F_u} \alpha_{uv} \boldsymbol{\varphi_{uv}}[o_u(i), o_v(t_v)] + \alpha_{u0} \boldsymbol{\theta_u^i}[o_u(i)] \tag{4.6}$$

Different from the traditional aggregation process, this formula redefines the opinion influence process from a temporal and interaction view. α_{uv} is the strength of influence that v exerts on u. α_{u0} represents u's degree of opinion stubbornness. α_{uv} and

α_{u0} could also be regarded as the probability that u chooses the chain v to determine her/his state. The sum of α_{uv} and α_{u0} is therefore one. If α_{u0} is up to 1, it means that u is difficult to be influenced by others.

Learning

The likelihood function of the opinion sequences for all users is written as Equation 4.7, and the learning problem becomes to maximize the likelihood function.

$$P(S) = \prod_{u} \prod_{i=2}^{n_u} (\sum_{v \in F_u} \alpha_{uv} \boldsymbol{\varphi_{uv}}[o_u(i), o_v(t_v)] + \alpha_{u0} \boldsymbol{\theta_u^i}[o_u(i)]) \tag{4.7}$$

The TIM model is characterized by three parameters $(\boldsymbol{A}, \boldsymbol{\Phi}, \boldsymbol{\Theta})$ where $\boldsymbol{A}$ is the influence strength parameter, $\boldsymbol{\Phi}$ is the state transition probability and $\boldsymbol{\Theta}$ represents personal opinion preference. Our objective is to maximize the likelihood function $P(\boldsymbol{S})$ by learning the three parameters. Following the temporal Markov assumption, $\boldsymbol{\Theta}$ which describes a user's prior opinion distribution could be estimated by the proportion of each sentiment in the previous posting history. Besides, each state transition probability matrix $\boldsymbol{\varphi_{uv}} \in \boldsymbol{\Phi}$ is independent of the other state transition probability matrices. Here, a Laplace smoothing technique is used. For each user pair u and v, $\boldsymbol{\varphi_{uv}}$ satisfies $\sum_o \boldsymbol{\varphi_{uv}}[o, o'] = 1$, for $o \in O$. Thus, $\boldsymbol{\varphi_{uv}}$ could be inferred by taking each connected chain of u and v alone according to the maximum likelihood estimate for a Markov chain [5]. The method is to count the frequency of each type of state transition observed in the sequences of u and v, and the state transition probability could be obtained via normalization.

Given the learned $\boldsymbol{\Theta}$ and $\boldsymbol{\Phi}$, the objective becomes to maximize $P(\boldsymbol{S})$ by inferring interpersonal influence $\boldsymbol{A}$. To facilitate the calculation, we construct the influence vector for each user u, $\boldsymbol{\alpha_{u\cdot}} = [\alpha_{u0}, \alpha_{u1}, \cdots, \alpha_{um_u}]^T$ to represent the degree of stubbornness and influence strength together. After rewriting the likelihood function to a log likelihood and removing the parts not related to $\boldsymbol{A}$, we observe that the influence vector for a user is independent of the influence vectors of others' vectors in the log likelihood function as shown in Equation 4.8.

$$\mathcal{L} = \sum_u \boldsymbol{\alpha_{u\cdot}} \log(\sum_v \alpha_{uv} \boldsymbol{\varphi_{uv}}[o_u(i), o_v(t_v)] + \alpha_{u0} \boldsymbol{\theta_u^i}[o_u(i)]) \tag{4.8}$$

s.t., for each $u \in V$, $\sum_v \alpha_{uv} + \alpha_{u0} = 1, \quad \alpha_{uv} > 0, \quad \alpha_{u0} > 0$ for $v \in F_u$.

As a result, we could update $\boldsymbol{\alpha}_{u\cdot}$ for each user individually. The objective function is simplified by separating it to per chain likelihood function as shown in Equation 4.9.

$$\boldsymbol{\alpha_{u\cdot}} = \arg\max_{\boldsymbol{\alpha_{u\cdot}}} \log(\sum_v \alpha_{uv} \boldsymbol{\varphi_{uv}}[o_u(i), o_v(t_v)] + \alpha_{u0} \boldsymbol{\theta_u^i}[o_u(i)]) \tag{4.9}$$

s.t., $\sum_v \alpha_{uv} + \alpha_{u0} = 1, \quad \alpha_{uv} > 0, \quad \alpha_{u0} > 0$ for $v \in F_u$.

By using Jensen's inequality [22], this per chain likelihood function is proved concave in $\boldsymbol{\alpha_{u\cdot}}$. To learn the optimal solutions of $\alpha_{u\cdot}$ from Equation 4.9, we remove the equality constraints by representing α_{u0} as $1 - \sum_v \alpha_{uv}$, where $v \in F_u$. Hence, all the constraints become inequality constraints as follows:

$$\alpha_{uv} > 0 \quad and \quad 1 - \sum_v \alpha_{uv} > 0$$

The problem becomes an optimization problem under the inequality constraints. A common solution to the inequality constrained optimization problem is the interior-point method [22, 127]. It approximates the objective function by adding the constraints with the logarithm barrier function, and forms the problem to an unconstrained optimization problem, which is presented in Equation 4.10, where β_1, β_2 represent approximating accuracy.

$$\max_{\boldsymbol{\alpha_3 u\cdot}} F = \log(\sum_v \alpha_{uv} \boldsymbol{\varphi_{uv}}[o_u(i), o_v(t_v)] + \alpha_{u0} \boldsymbol{\theta_u^i}[o_u(i)]) + \beta_1 \log(1 - \sum_v \alpha_{uv}) + \beta_2 \sum_v \log(\alpha_{uv}) \tag{4.10}$$

Algorithm 4.1 Path-following Method

Input:

opinion sequences $\{o_u(1), o_u(2), \cdots, o_u(n_u)\}$ for each user and the posting time is $t_u(i)$.
influencing neighbor set F_u for each u
personal opinion distribution $\mathbf{\Phi}$
state transition probability $\mathbf{\Theta}$
tolerance ϵ, updating step μ and maximal number of iterations I, approximating accuracy β_1, β_2, learning rate γ.

Initialization:

```
initialize the influence strength with equal value α_uv = 1/(|F_u|+1) for v ∈ {F_u ∪ {0}}
for u=1 to N do: do
    while β_1 > ε do: do
        for iter=1 to I do: do
            Compute α*_u· according to Equation 4.12
            if Convergence then
                break
            end if
        end for
        update α_u· = α*_u·, β_2 = β_1 = β_1/μ
    end while
end for
return
α_uv and α_u0 = 1 − Σ_v α_uv
```

Because the solution of the optimization problem in Equation 4.10 is an approximated solution, we use the path-following method presented in Algorithm 4.1 [22] to make the approximated solution close to the original solution. It iteratively solves the unconstrained optimization problem in Equation 4.10 by decreasing the values of β_1 and β_2. In each iteration, we apply the gradient descent method to find the optimal solution of the unconstrained problem. The starting point for the current iteration is the optimized point found in the last iteration. The updating rule is

$$\boldsymbol{\alpha}^{\star}_{\boldsymbol{u}\cdot} = \boldsymbol{\alpha}_{\boldsymbol{u}\cdot} + \gamma \frac{\partial F}{\partial \boldsymbol{\alpha}_{\boldsymbol{u}\cdot}} \tag{4.11}$$

For the v-th element $\frac{\partial F}{\partial \alpha_{uv}}$ of the derivative $\frac{\partial F}{\partial \boldsymbol{\alpha}_{\boldsymbol{u}\cdot}}$ is

$$\frac{\partial F}{\partial \alpha_{uv}} = \sum_{i=2}^{n_u} \Big(\frac{\boldsymbol{\varphi}_{\boldsymbol{uv}}[o_u(i), o_k(t_v)] - \alpha_{uu} \boldsymbol{\theta}^{\boldsymbol{i}}_{\boldsymbol{u}}[o_u(i)]}{(\sum_v \alpha_{uv} \boldsymbol{\varphi}_{\boldsymbol{uv}}[o_u(i), o_v(t_v)] + (1 - \sum_v \alpha_{uv}) \boldsymbol{\theta}^{\boldsymbol{i}}_{\boldsymbol{u}}[o_u(i)]} \Big) - \beta_1 \frac{1}{\log(1 - \sum_v \alpha_{uv})} + \beta_2 \frac{1}{\sum_v \log(\alpha_{uv})} \tag{4.12}$$

Note that the influence parameter $\boldsymbol{\alpha}_{\boldsymbol{u}\cdot}$ can be updated individually for each user. It provides us the opportunity to parallelize the learning process. With more CPU cores, the multi-parallel task could speed up.

4.5 Experiments and Discussions

4.5.1 Experimental Set-up

We conduct experiments on the dataset of popular products, i.e., “Samsung Galaxy”, “Xbox” and “PlayStation”. The data collection method and the statistics of dataset have been introduced in Chapter 3. For each topic, we split the collected data into the training set and test set. The first 90% of tweets of each user are used for training and the remaining 10% for testing. The statistics of the dataset is presented in Table 4.2.

Table 4.2: Dataset statistics.

Method		negative opinion	neutral opinion	positive opinion	# of instances
Samsung Galaxy	training set	11.15%	40.43%	48.42%	696916
	test set	10.31%	38.58%	38.58%	81997
Xbox	training set	16.26%	41.61%	42.13%	244787
	test set	17.04%	41.21%	41.75%	29444
PlayStation	training set	11.86%	63.32%	24.82%	704238
	test set	13.12%	59.31%	27.57%	80650

To evaluate the performances of sentiment prediction, the metric we use is the prediction accuracy, which is defined as the accuracy of correctly predicted test instances.

$$Accuracy = \frac{the\ number\ of\ correctly\ predicted\ tweets}{the\ number\ of\ tweets\ in\ the\ testing\ set}$$

For a more comprehensive analysis, we further evaluate the ability of all influence models on predicting different sentiments. The F1 score considers both precision and recall. It is used as the measurement on each sentiment category, including *F1_Pos* for positive sentiment, *F1_Neu* for neutral sentiment, and *F1_Neg* for negative sentiment.

4.5.2 Compared Methods

We testify TIM on its predictive ability for a user's future sentiment. To predict u's sentiment, we refer to Equation 4.6. The performance of TIM is compared with the results of one individual-based model, i.e., temporal Markov model (TMM), and four influence-based models including Degroot, Flocking, ASLM and Voter models.

TMM (Temporal Markov Model): It assumes that an individual's next sentiment is only decided by her/his most frequent sentiment observed in the past. The assumption is consistent with the temporal Markov assumption proposed in Section 4.4. Formally,

$$\hat{o}_u(i) = \arg\max_o \boldsymbol{\theta_u^i}$$

Degroot: It is a classical opinion influence model that describes opinion influence as an aggregation process. It assumes that a user takes the averaged sentiments of her/his neighbors as her/his new sentiment [46]. Different from TIM that characterizes opinion as discrete sentiment polarity, the Degroot model represents opinion as a continuous score. At each timestamp $t_u(i)$, the updated opinion $s_u(i)$ is the average of the sentiment scores of her/his neighbors, which is denoted as:

$$s_u(i) = \alpha_{u0} s_u(i-1) + \sum_{v \in F_u} \alpha_{uv} s_v(t_v)$$

where $s_u(i-1)$ represents the sentiment score of u's previous tweet and $s_v(t_v)$ represents the most recently published sentiment of u's neighbor v before $t_u(i)$. α_{u0} and α_{uv} represent u's stubbornness and interpersonal influence between u and v, respectively, which are similar to the parameter $\boldsymbol{A}$ in TIM. Similar to TIM, the parameter $\boldsymbol{\alpha_{u\cdot}}$ satisfies $\sum_{v \in F_u} \alpha_{uv} + \alpha_{u0} = 1$ and $\alpha_{uv} > 0, \alpha_{u0} > 0$. The sentiment scores used in the Degroot model are obtained with the opinion processing method presented in Chapter 3. Given the predicted sentiment score, we could infer the sentiment polarity, i.e., a positive sentiment value indicates positive opinion, a negative sentiment indicates a negative opinion and the zero value indicates the neutral opinion.

Flocking: It is a variant of the Degroot model with the assumption that a user u only trusts those neighbors who have similar opinions with her/himself [75]. Given the past sentiments of all users, we construct the trusted neighbor set $F_u(trust)$ for each user u.

$$F_u(trust) = \{||\overline{s_u} - \overline{s_v}| \leqslant \varepsilon \ and \quad v \in F_u\}$$

where $\overline{s_u}$ and $\overline{s_v}$ represent the averaged sentiment score of user u and her/his neighbor

v. A user u takes the averaged sentiment of her/his trusted neighbors, i.e.,

$$s_u(i) = \alpha_{u0} s_u(i-1) + \sum_{v \in F_{u(trust)}} \alpha_{uv} s_v(t_v)$$

AsLM: It is the state-of-the-art opinion influence model proposed by [44], which has been demonstrated to be superior to other opinion influence models on sentiment prediction. It has the assumption similar to Degroot except that the influence strength α_{u0} and α_{uv} have no boundary and are allowed to be positive, negative or zero.

Voter: Different from previous three methods that represent opinions with continuous values of sentiment, the Voter model [40, 77], like our TIM, represents opinion with discrete sentiment polarity. We choose a variant of the voter model that follows the majority rule [97]. In this method, at each timestamp $t_u(i)$, u selects the major sentiment occurring in the set $\{o_u(i-1), o_1(t_1), \cdots, o_v(t_v), \cdots\}$, where $v \in F_u$ and $t_u(i-1) < t_v < t_u(i)$.

4.5.3 Overall Performances

The experimental results are reported in Table 4.3. Over the three topics, TIM consistently achieves better performances than both individual-based method and opinion influence models in sentiment prediction. We have the following findings.

First, we find that TIM achieves better results compared with state-of-the-art opinion influence models in terms of all evaluation metrics. Most of the theoretical opinion influence models simplify the opinion influence process as an aggregation process. However, they lack the ability to uncover the temporal properties of user interactions. The better performances of TIM prove that temporal interactions are important for understanding the opinion influence process and are well captured by the proposed model.

Table 4.3: Performances on three products.

(a) Samsung Galaxy

Method	Accuracy	F1_Neg	F1_Neu	F1_Pos
TMM	0.6408	0.1341	0.7153	0.5998
Degroot	0.5772	0.1688	0.6857	0.4408
Flocking	0.5819	**0.2414**	0.6939	0.3950
AsLM	0.5481	0.1555	0.6826	0.4597
Voter	0.5822	0.1773	0.6593	0.5425
TIM	**0.6576**	0.1683	**0.7276**	**0.6305**

(b) Xbox

Method	Accuracy	F1_Neg	F1_Neu	F1_Pos
TMM	0.5591	0.1875	0.6124	0.5699
Degroot	0.4893	0.1801	0.5897	0.4265
Flocking	0.4407	0.0861	0.5621	0.2298
AsLM	0.4935	0.1963	0.7233	0.3056
Voter	0.5116	0.1861	0.5471	0.5308
TIM	**0.5712**	**0.2349**	**0.6390**	**0.5858**

(c) PlayStation

Method	Accuracy	F1_Neg	F1_Neu	F1_Pos
TMM	0.6526	0.1661	0.7716	0.4551
Degroot	0.6115	0.0701	0.7510	0.2023
Flocking	0.2023	0.1037	0.7332	0.1789
AsLM	0.5707	0.1985	0.7216	0.3571
Voter	0.5443	**0.2481**	0.6543	0.4603
TIM	**0.6667**	0.2205	**0.7811**	**0.4705**

Second, we find that the performances of the four influence-based models vary a lot in the three topics, and it is hard to say which model performs the best on all the evaluation metrics. The existing opinion influence models are developed according to their assumption on the final state of a dynamic system. They cannot be generalized to model opinion influence on different topics. However, our proposed model achieves a balanced performance on all evaluation metrics.

Third, we also find that TIM outperforms the individual-based model TMM in terms of all evaluation metrics on three products. It demonstrates that opinion influence actually exists during a user's decision making process. Opinion influence is well characterized by the two interaction-based opinion influence indicators.

Finally, we also find that though the individual-based TMM achieves good result on the overall prediction accuracy, its prediction ability on the negative sentiment prediction is bad compared with the influence-based methods. It demonstrates that opinion influence plays an important role in getting people to form negative impressions on the brand or product. As told in the consumer studies [118], the greater importance is given to the negative information than to positive information in the general evaluation of a given product/brand. It is very important for companies to manage their negative images, which is better foretasted by the opinion influence models.

4.5.4 Evaluations on Interpersonal Influence

Most existing influence studies focused on measuring the global influence of each user, but the interpersonal influence between each user pair has not been fully explored. After modeling opinion influence with TIM, interpersonal influence is captured and formulated by the parameter $\boldsymbol{A}$. In this section, we conduct an experiment to further evaluate the learned interpersonal influence. Because interpersonal influence is difficult to be evaluated directly, we still rely on the task of future sentiment prediction. We develop two variants of TIM, which use different approaches to measure interpersonal influence. TIM is compared with these two variants to verify the effectiveness of the learned interpersonal influence.

TIM_S1: It assumes that each user has the same influence on all her/his followers, which is popularly used in social media analysis [170]. This assumption is

Table 4.4: Evaluations on interpersonal influence .

(a) Samsung Galaxy

Method	Accuracy	F1_Neg	F1_Neu	F1_Pos
TIM_S1	0.6292	0.1630	0.7012	0.6066
TIM_S2	0.6454	0.1569	0.7164	0.6224
TIM	**0.6576**	**0.1683**	**0.7276**	**0.6305**

(b) Xbox

Method	Accuracy	F1_Neg	F1_Neu	F1_Pos
TIM_S1	0.5452	0.2380	0.6198	0.5547
TIM_S2	0.5565	0.2196	0.6263	0.5709
TIM	**0.5712**	**0.2349**	**0.6390**	**0.5858**

(c) PlayStation

Method	Accuracy	F1_Neg	F1_Neu	F1_Pos
TIM_S1	0.6414	0.2153	0.7580	0.4552
TIM_S2	0.6507	0.2042	0.7655	0.4556
TIM	**0.6667**	**0.2205**	**0.7811**	**0.4705**

basically in accordance with the studies that measure the overall social influence for every user. We employ the PageRank algorithm [129] to learn the influence power of each individual user u. For each u, the normalized influence power of u and u's neighbors v are defined as the influence strength α_{u0} and α_{uv}.

TIM_S2: It is the other variant of TIM. Different from TIM_S1, TIM_S2 assumes that each user u is equally influenced by all his/her friends. It is also a popularly used assumption in social influence analysis [142]. In TIM_S2, $\alpha_{uv} = \alpha_{u0} = \frac{1}{|F_u|+1}$ for $v \in F_u$.

The experimental results are presented in Table 4.4. TIM_S1 that calculates users' influence power based on the network structure is a bit worse than TIM_S2 that has the equal influence assumption. This suggests the shortcoming of judging influence only from the network structure without considering the effects of influence in real online communication. Compared with both TIM_S1 and TIM_S2, TIM that captures interpersonal influence from communication traces performs the best. It demonstrates the reasonability of our influence assumption, and the effectiveness of TIM in learning interpersonal influence.

4.6 Chapter Summary

In this chapter, we address the issue of learning opinion influence from users' temporal interactions. A framework is developed by temporally correlating the opinion dynamics of network users. Furthermore, two assumptions specific to social interactions are integrated into the framework, which are the temporal Markov assumption and interaction-based opinion influence assumption. The experiment results demonstrate the effectiveness of the temporal opinion model, especially in uncovering temporal dynamics of user interactions.

Chapter 5

Content-based Opinion Influence Modeling

5.1 Chapter Overview

In Chapter 4, opinion influence is studied from a dynamic point of view, where opinion influence is characterized as the correlation between users' opinion dynamics during the long-term interactions. We have proved that understanding the properties of temporal interactions is of great importance for opinion influence estimation. However, there exist many communication rounds during the long-term social interactions, and a comprehensive understanding of each communication round remains incomplete. In this chapter, we take a closer look at the important elements in each communication round, and explore their effects on opinion influence modeling.

As stated in the area of business communication[1], one communication round is usually composed of two important elements as shown in Figure 5.1. They are the message, which is the transmitted information during the communication and the involved users who possess different personalities and identities. A communication round can be described as follows: a sender writes a message of information, and transmits it to a receiver. After the receiver receives the information, s/he will give a

[1] http://managementstudyguide.com/components-of-communication-process.htm

Figure 5.1: Two elements in the communication round

response. Considering opinion influence is the result of communication, understanding the communication process would benefit modeling of opinion influence.

The main purpose of communication is to transmit messages, which contains the information a person would like to convey to her/his audience. In existing studies on opinion influence, a message is summarized as the sentiment polarity, but the textual content that truly reflects the information a user receives from her/his neighbors is totally ignored. We are the first to explore the effects of the content information in social interactions and utilize it to model opinion influence process.

Besides, the involved users including senders and receivers are the other important element in a communication round. A sender is a person who makes use of the symbols (e.g., words) to convey the message to other users, and a receiver is a person who reads the message from a specific sender and gives the feedback. On social media, users possess different personalities and each kind of personality has its own effect on communication. For example, a sender, as an expert, who states professional reviews on a product, would have a higher possibility to change receiver's opinions. A receiver, as a fan towards a product who has a strong belief in the product is difficult to be influenced by other users. Thus, understanding users' personalities and their identities help us to understand their behaviors during the communication and further to capture opinion influence.

In the following sections, we start by a neural opinion influence model, which is the first to utilize the advances of neural networks to understand opinion influence from the textual content (Section 5.2). Furthermore, based on the neural opinion influence model, we show how to incorporate the different identities of participants into opinion influence modeling (Section 5.3).

5.2 Neural Opinion Influence Model

5.2.1 Introduction

On most social media platforms, people exchange their views by posting and replying through textual messages. For example, Epinion, a review website, allows people to write their own experiences of purchases; Quora, a community Q&A platform, offers opportunities for people to ask questions in natural language and people can also answer the questions posted by others by writing their point of views. The User-generated-content (UGC) created by users appears as an important component during their communication, and it provides fruitful information to capture opinion influence between users and explains how users form or change their opinions.

The existing influence models including the temporal opinion influence model proposed in Chapter 4 characterize opinions in the form of either discrete categories of sentiment polarities, such as positive, negative and neutral sentiments [75, 58, 44], or continuous scales of opinion strengths [40, 46, 28, 179]. However, the summarized opinion states cannot effectively reflect the reason why a user changes her/his opinion when the content information is totally ignored. Even if two messages have the same sentiment, different content information may result in different effects on others' opinions. We take two postings about the product "Samsung Galaxy" as examples.

(1): *I can't post gifs on this stupid Galaxy S6.*

(2): *Just lost my new Galaxy S6 and very sad.*

Here, (1) expresses a complaint to a problem related to the picture-posting function provided by the product, which may make other users have an unfavorable impression on the product. On the other hand, (2) simply expresses the personal feeling of sadness and regret due to the loss of the product. There is no bad effect on "Samsung Galaxy" transmitted through this message at all. The above two examples show that the summarized sentiment of the opinion, i.e., the negative sentiment in these cases, is not able to differentiate the opinion effects of different content on other users. Hence, it is necessary to well understand the role of content information in the opinion influence process.

Hence, the problem of opinion influence modeling becomes discovering the underlying relevance between a person's opinion and the received content information from her/his neighbors. The intuitive solution is to employ the co-occurrence patterns of one's opinion and her/his neighboring messages. However, as the data grows, the patterns of co-occurrences can be sparse and ineffective for prediction. Vector representations of words and phrases have been successfully applied in many Natural Language Processing (NLP) tasks [13, 100, 37]. By encoding the semantic information, word embedding makes it possible to overcome the problem of "curse of dimensionality".

Therefore, we propose a Neural Opinion Influence Model (NIM) by vectorizing the discrete words with continuous vectors. As far as we know, we are the first to employ the Neural Network (NN) advances in opinion influence modeling. In NIM, we focus on the textual messages exchanged in each communication round and follow the assumption used in existing opinion influence models, which only considers the most recently received information for prediction. Each opinion word is represented as a dense vector in the continuous space. We then compose the opinion word vectors of one's previous message and her/his neighboring messages to form the social opinion context vector and feed the vector to a softmax layer for sentiment prediction. Two social relation factors including stubbornness and interpersonal influence are integrated to construct the social opinion context vector. The proposed neural opinion influence model bears similarities with the neural language model [13]. In the scenario of the opinion influence process, we regard the neighboring opinions and one's previous opinion as the "contexts", and the "target" is one's future opinion sentiment. The model has a more complex framework since the factors including stubbornness and interpersonal influence are considered together with word embeddings.

Different from previous state-based opinion influence models, NIM considers the textual information included in opinions, which better depicts the opinion influence process. In the experiments conducted on Twitter datasets, NIM performs better than other state-of-the-art opinion influence models. Besides, the analysis of the users' expressions with different influence powers increases the interpretability of the opinion influence model. It could provide references for companies to understand the different effects of different wordings in order to better manage their social accounts.

5.2.2 Related Work in Word Representation Learning

Word representation learning is to represent an individual word w_i with a vector $\boldsymbol{v_i}$. A straightforward approach is one-hot representation [64], which uses an N-dimension vector to represent the word w_i, and N is the size of the lexicon. The ith element of the vector is one and the other elements are zero. However, the simple one-hot representation is sparse when the size of the word lexicon increases. Besides, only the index of the word is included in the one-hot representation. The semantic information of the word is not captured.

In recent years, learning word representations with neural networks achieves great performances on a variety of natural language processing tasks, e.g., machine translation [9, 36, 110], dialogue generation [147, 78, 105], question answering [176, 173] and etc. The idea of word representation was first proposed by Bengio et al., with a probabilistic neural language model [13]. The model simultaneously learned the distributed representation for each word along with the joint probability function of a word sequence according to the learned word representations. Given the contextual words of a specific word w_i, the model first embeded all contextual words into their dense representations by looking up into an embedding matrix. The contextual word vectors were concatenated and then fed into a feed-forward neural network. The probability of the word w_i occurring along with the contextual words was then predicted by a softmax layer. The embedding matrix and the parameters of the neural network were learned with a back-propagation method. The learned word representations provided an effective way to sufficiently capture the semantic meaning of words and to cope with the "curse of dimensionality" problem.

Later, based on the neural language model, several techniques were developed to improve the efficiency of word representation learning using large-scale corpora. Bengio and Secenal [14] proposed a new learning technique based on importance sampling, which optimized the model with the observed positive examples and the sampled negative examples. Considering the large dimension of the lexicon, Morin [119] proposed a hierarchical softmax function to effectively reduce the number of parameters in the output layer. In 2013, Mikolov et al. [116, 115] introduced two popular frameworks for learning word representation, including the continuous bag-of-words (CBOW) and continuous skip-gram. The CBOW model predicted the current word based on the embeddings of its contextual words, while the skip-gram model predicted the contextual words given the embedding of the current word. The two approaches were later released to the public in a widely-used word2vec toolkit for academia and industry use.

A lot of researchers also focused on utilizing different kinds of techniques to capture richer semantic information. For example, Levy and Goldberg incorporated the syntactic contexts that were derived from a dependency parse-tree [103] into the word embedding model. Li and Jurafsky proposed a model to learn multi-sense of word embeddings [104]. Recently, more and more studies were developed to learn the task-specific representations, such as word embedding for sentiment classification [155], or for document classification [177]. The learned word representations heavily relied on the application where it was used. Our work focuses on learning the opinion-specific word representations tailored for sentiment prediction on social media.

The learned word representations provide the opportunities for researchers to obtain the phrase-level and sentiment level representations by composing all vectors of words in a phrase or a sentence. A basic composition method is to use the weighted average of all word vectors [180, 116]. In [116], they used a simple data-driven approach, where the phrases were formed based on the unigram and bigram counts of the words. Further considering the syntactic structure of phrases or sentences, a method combining words according to their dependencies in the syntactic tree was proposed [149]. Different from the existing composition method, which only considers the semantic and syntactic structure, we propose a social composition method. It combines individual opinions to form an opinion context by considering two social-related factors, including stubbornness and interpersonal influence.

Table 5.1: Definition of notations

Notation	**Description**
V	a set of $\|V\| = N$ users.
F_u	a set of u's neighbors in the network.
$S_u = \{W_u(1), \cdots, W_u(n_u)\}$	a sequence of u's posting records.
$W_u(i)$	opinion words included in u's i-th message
$o_u(i)$	the sentiment of u's i-th message.
$t_u(i)$	the time u publishes the i-th message.
$C_u(i) = \{C_u^1(i), \cdots, C_u^v(i), \cdots, C_u^{m_u}(i)\}$	neighboring opinions u takes as the reference to change her/his future opinion

5.2.3 Problem Formulation

The notations used in this section are summarized in Table 5.1. For each user $u \in V$, her/his posting message at time stamp $t_u(i)$ is represented by the set $W_u(i) = \{W_{u,1}(i), \cdots, W_{u,|W_u(i)|}(i)\}$, which contains all the opinions words included in the message. All messages posted by u constitute a posting sequence $S_u = \{W_u(1), \cdots, W_u(n_u)\}$. Considering the information from one's neighbors is an important factor for a user to change her/his opinions, we construct the set $C_u(i)$ to include the information that triggers u to change her/his next opinion. It contains the most recent opinions u receives from each neighbor F_u^v since $t_u(i-1)$, which is denoted by $W_{F_u^v}(t_v)$. Here t_v satisfies that $t_u(i-1) < t_v < t_u(i)$. For brevity, we write the set as $C_u(i) = \{C_u^1(i), \cdots, C_u^v(i), \cdots, C_u^{m_u}(i)\}$. Each element $C_u^v(i)$ is a set of opinion words included in the message $W_{F_u^v}(t_v)$ and is represented by $C_u^v(i) = \{C_{u,1}^v(i), \cdots, C_{u,|C_u^v(i)|}^v(i)\}$. If there exists no posting from neighbor v during the time period, $C_u^v(i)$ is an empty set.

We assume that a user's opinion at a particular timestamp is determined by the new information s/he receives from neighbors and the opinion s/he holds before. The problem of sentiment prediction is defined as follows. Given the neighboring opinion information u receives in previous timestamp $C_u(i)$ and her/his previous personal opinion $W_u(i-1)$, the objective is to uncover opinion influence between users and use it to predict the future sentiment $o_u(i)$ of u at the timestamp $t_u(i)$.

5.2.4 Proposed Model

We propose a novel neural opinion influence model relying on representation learning to solve the sentiment prediction problem. The model is based on the feed-forward neural network. It starts with a word embedding layer, where each opinion word

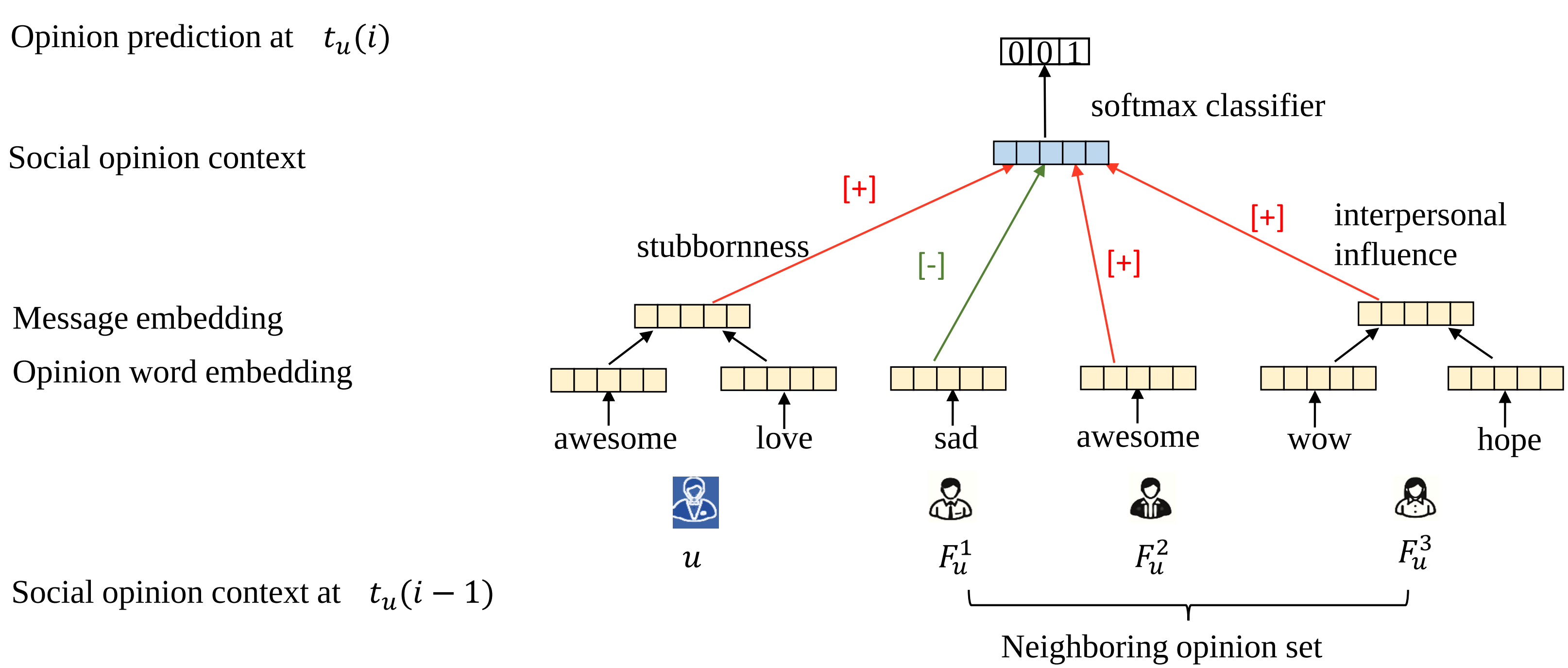


Figure 5.2: The graphical representation of the NIM on the opinion prediction

is mapped to a dense vector. A hidden layer follows with a composition layer to construct the social opinion context vector by concerning the social relation factors. Afterwards, the social opinion context vector is used to predict the user's future sentiment in the output layer. The graphical illustration of the proposed model is shown in Figure 5.2.

Word Embedding Layer

An intuitive solution to predict the future sentiment is to uncover the correlation between the next sentiment and the combination of the received opinion information including the personal prior message $W_u(i-1)$ as well as the neighboring opinion sequence $C_u(i)$. With the growth of the vocabulary of opinion words, the number of possible combinations of the discrete opinion words grows exponentially and many of them may not be observed in the dataset. A fundamental problem called "curse of dimensionality" will happen [12]. Conventional classifiers such as logistic regression [3], SVM classification [151] and etc. will unavoidably encounter the overfitting problem when modeling the high dimensional data and the classification performance will be harmed. To overcome the challenge raised by the curse of dimensionality, the dimensionality reduction is required. Recently, word embedding achieves great performances in natural language processing tasks [13, 116] not only because of its ability to reduce the dimensionality of word vocabulary, but also due to its ability to preserve the semantic information. The word embedding provides a convenient way to represent the joint distribution of discrete words by mapping the discrete opinion words to the dense and continuous vectors. Thus, we resort to the word embedding techniques to capture the role of content information in opinion influence modeling. We convert the bag-of-word representation W_j into a dense and low-dimensional vector $\boldsymbol{w_j}$ by an embedding projection matrix $\boldsymbol{\Phi} \in \mathbb{R}^{d_w \times v}$, where v is the total number of opinion words.

$$\boldsymbol{w_j} = \boldsymbol{\Phi} W_j \tag{5.1}$$

Given the embedding matrix $\boldsymbol{\Phi}$, the personal prior opinion $W_u(i-1)$ can be easily represented by the vector $\boldsymbol{p_u}(i-1)$.

$$\boldsymbol{p_u}(i-1) = \sum_{j=1}^{|W_u(i)|} \boldsymbol{\Phi} W_{u,j}(i-1) \tag{5.2}$$

where $\boldsymbol{p_u}(i-1) \in \mathbb{R}^{d_w}$.

Similarly, each element $C_u^v(i)$ included in the neighboring opinion set $C_u(i)$ is also converted to the summation over the representations of all opinion words, which is denoted by $\boldsymbol{s_u^v}(i)$.

$$\boldsymbol{s_u^v}(i) = \sum_{j=1}^{|C_u^v(i)|} \boldsymbol{\Phi} C_{u,j}^v(i) \tag{5.3}$$

Social Opinion Context Composition

At each timestamp $t_u(i)$, user u holds a prior opinion and decides to update her/his opinion towards a specific topic after receiving the opinions from her/his neighbors. After representing the messages with dense vectors, the information that affects a user's future opinion can be formulated as the composition over the personal prior opinion vector $\boldsymbol{p_u}(i-1)$ and the neighboring opinion vector $\boldsymbol{s_u^v}(i)$.

Traditional composition methods form the phrase or sentence vector by combining word vectors with the weights obtained from the data, or applying the matrix transformation to the concatenation of word vectors [100]. In this work, we propose a composition method utilizing two social relation factors that have been commonly considered in previous influence models [43, 44]. The social relation factors include stubbornness and interpersonal influence between users. The stubbornness factor describes how much a person insists on her/his previous opinion, and the interpersonal influence factor represents the strength of influence a neighbor exerts on the user. Considering interpersonal influence has the linear property [44], the model averages the personal prior opinion vector $\boldsymbol{p_u}(i-1)$ and the neighboring opinion vector $\boldsymbol{s_u^v}(i)$ with the social relation factors. Formally, it is denoted as follows.

$$\boldsymbol{c_u}(i) = tanh(\alpha_{u0})\boldsymbol{p_u}(i-1) + \sum_{v=1}^{m_u} tanh(\alpha_{uv})\boldsymbol{s_u^v}(i) \tag{5.4}$$

where α_{uv} represents the interpersonal influence on user u's opinion from the vth neighbor, and α_{u0} represents u's stubbornness.

Note that in the previous study, the social relation factors $\boldsymbol{\alpha}_{u.}$ are assumed to represent the probabilities that a person selects the opinions from her/his neighbors. Thus, its value should be between 0 and 1. In this study, the proposed framework based on neural network is more flexible to consider opinion influence. We allow opinion influence be either positive or negative. The idea of polarity-related influence was firstly proposed by [44], and was proved quite effective for sentiment prediction on social network. The positive influence happens when a user trusts her/his friend and s/he will accept the opinion of her/his friend and express the same one. The negative influence implies that a user gets influenced by her/his friend, but to the opposite direction. Thus, the two social relation factors are limited between -1 and 1 by using a tangent function in Equation 5.5, which allows both positive and negative influence.

$$tanh(\alpha_{uv}) = \frac{e^{\alpha_{uv}} - e^{-\alpha_{uv}}}{e^{\alpha_{uv}} + e^{-\alpha_{uv}}} \tag{5.5}$$

Sentiment Prediction

Finally, the opinion context vector $\boldsymbol{c_u}(i)$ is taken as the features to predict the future sentiment in the output layer. The output layer of our approach is expressed by the following equation.

$$P(o_u(i)|\boldsymbol{c}_{\boldsymbol{u}}(i)) = \sigma(\boldsymbol{V}\boldsymbol{c}_{\boldsymbol{u}}(i) + \boldsymbol{b}) \tag{5.6}$$

σ is a softmax function, which represents the probability of z belonging to the jth class.

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{k=O} e^{z_k}} \tag{5.7}$$

where $\boldsymbol{V} \in \mathbb{R}^{O \times d_w}$, and $b \in \mathbb{R}^O$. O is the number of sentiment, and it is set 3 in the study.

Learning

The model is parameterized by the social relation factors $\boldsymbol{\alpha}$, the word embedding matrix $\boldsymbol{\Phi}$, and the output parameters $\boldsymbol{V}, \boldsymbol{b}$. The objective function to be maximized is the log-likelihood of all opinion behavior sequences defined in Equation 5.8.

$$\mathcal{L}(O) = \sum_{u=1}^{N} \sum_{i=1}^{n_u} \log P(o_u(i)|C_u(i), W_u(i-1)) \tag{5.8}$$

We learn the model parameters using the stochastic gradient decent (SGD) algorithm. The dimensionality of the word embedding d_w is set to 30. During the training phrase, we normalize the gradients if the norm exceeds 1 [134]. The training phrase stops when the training error has a decrease less than 1 or reaches the maximum iteration length of 100. The model is implemented by the Theano library [10].

5.2.5 Experiments and Discussions

Experimental Set-up

To demonstrate the effectiveness of exploring content information in opinion influence modeling, we compare NIM with the state-based opinion influence models, including

Degroot, Flocking, Voter and AsLM, which have been introduced in Section 4.5.2 in details. To further verify the effectiveness of the word representation and the proposed neural network framework, we develop another two content-based opinion influence models. One is **Content_SVM**, which is implemented with LIBSVM [27]. The model trains the SVM classifier individually for each user by taking all the neighboring opinion words and the opinion words in one's previous tweet as the features. Because SVM does not have the ability to learn the word representations, the one-hot representation of each word is used as the features. To be consistent with the linear influence assumption, the linear kernel is used in SVM training process. The other content-based model **NIM_noEmb** is a variant of NIM, which employs the same framework as NIM, but utilizes the one-hot vector to represent each opinion word instead of the dense word embedding. The parameters of each model are set for their best performances experimentally.

We follow the same way to split the training and test sets as described in Chapter 4. The prediction accuracy and the F1 score on three sentiment polarities are used as the evaluation metrics.

Performance Evaluation

The results are presented in Table 5.2. In short, we observe the following findings.

First, we find that NIM and its variant NIM_noEmb perform better than all the compared methods in all three topics and on almost all evaluation metrics. Specifically, the improvements compared with the best competitors on the positive sentiment prediction are 17.2% for the topic "Samsung Galaxy", 13.1% for the topic "Xbox" and 11.2% for the topic "Samsung Galaxy". The results verify that the content information actually provides a better way to understand the opinion influence process, and our proposed Neural Network framework provides an effective way to understand the content information, as well as capture the opinion influence

Table 5.2: Performances on three products.

(a) Samsung Galaxy

Method	Accuracy	F1_Neg	F1_Neu	F1_Pos
Degroot	0.5772	0.1688	0.6857	0.4408
Flocking	0.5819	**0.2414**	0.6939	0.3950
AsLM	0.5481	0.1555	0.6826	0.4597
Voter	0.5822	0.1773	0.6593	0.5425
Content_SVM	0.5771	0.1436	0.6732	0.4918
NIM_noEmb	0.6434	0.1392	0.7243	0.6310
NIM	**0.6590**	0.2075	**0.7306**	**0.6357**

(b) Xbox

Method	Accuracy	F1_Neg	F1_Neu	F1_Pos
Degroot	0.4893	0.1801	0.5897	0.4265
Flocking	0.4407	0.0861	0.5621	0.2298
AsLM	0.4935	0.1963	**0.7233**	0.3056
Voter	0.5116	0.1861	0.5471	0.5308
Content_SVM	0.5586	0.2004	0.6106	0.5972
NIM_noEmb	0.5528	0.2143	0.6127	0.5783
NIM	**0.5694**	**0.2346**	0.6272	**0.6002**

(c) PlayStation

Method	Accuracy	F1_Neg	F1_Neu	F1_Pos
Degroot	0.6115	0.0701	0.7510	0.2023
Flocking	0.2023	0.1037	0.7332	0.1789
AsLM	0.5707	0.1985	0.7216	0.3571
Voter	0.5443	**0.2481**	0.6543	0.4603
Content_SVM	0.5556	0.2010	0.7458	0.4616
NIM_noEmb	0.6554	0.1233	0.7730	0.5106
NIM	**0.6653**	0.1301	**0.7813**	**0.5136**

Second, we also note that NIM performs much better than both content-based methods including Content_SVM and NIM_noEmb. Though the content information is considered in both methods, the content information is represented by the bag-of-word representation. Without considering the correlations among different words, the bag-of-word representation is ineffective to capture the semantic information. The superior performance of NIM demonstrates the power of word embedding in using the semantic information and modeling opinion influence. Besides, we also find that NIM_noEmb achieves better results compared with another content-based model Content_SVM, which uses the same features as NIM_noEmb. The better performances demonstrate the effectiveness of the neural network framework in modeling opinion influence.

Third, we also observe that though NIM gains a significant improvement on the prediction of the positive sentiment but a bit weak on the prediction of negative opinions. According to Table 4.2, the percentage of tweets in each dataset is less than 20%, and even about 10% for the topic "Samsung Galaxy". Because there are only a small amount of messages that express the negative opinions about the product, it maybe difficult for NIM to sufficiently learn the mechanism about how negative opinion formation.

Analysis of Wording of Influential Users

With the learned model, the companies could get the insights into how to become an influential voice on the social media by improving their wordings. We analyze different expressions used by users with different social opinion influence degrees in the network. Based on the learned interpersonal influences $\boldsymbol{\alpha}$, we calculate the influence strengths of Twitter users by averaging their outgoing influence strengths on their followers.

$$INF_v = \frac{\sum_u \alpha_{uv}}{N_u} \tag{5.9}$$

where $v \in F_u$ for all $u \in V$, and N_u is the number of users who follow v.

Based on the influence strengths INF_v, we divide users into three groups. The users with influence strengths more than 0.5 are categorized as the positively influential users. The users with influence strengths less than -0.5 represent the negatively influential users. The remaining users are regarded as the ordinary users with little influence. Finally, users with strong positive influence possess 5% among all users, and users with strong negative influence possess 2%.

We then extract the highly frequent opinion words from the users in the groups with positive influence and negative influence. The results show that the positively influential users more likely utilize the words describing the facts, e.g., "security", "special" and "impress". However, the tweets posted by strong negatively influential users are more emotional with the words like "Woo", "Wow" or the emoticons "o_o". The analysis indicates that the detailed information about the products tends to make positive effects, while heavily emotional expressions may annoy people and influence them to the opposite direction.

5.2.6 Conclusions

In this section, we report to characterize the users' tweets with detailed opinion content instead of discrete sentiment polarities. To the best of our knowledge, this is the first attempt to incorporate the content information into opinion influence modeling. The proposed model based on the feed-forward neural network framework is capable of learning the opinion word representations, which encode the semantic information of the opinions words, and learning interpersonal influence from opinion behaviors of all users. The experiments conducted on the Twitter datasets demonstrate the effectiveness of our proposed model on sentiment prediction. We also analyze the wording of the users with different influence powers. It certainly provides grateful insights for the companies to manage their account on social media. In the next section, we will further explore the effects of content information in opinion influence modeling by completing the details of communication process with users' identities.

5.3 Dual-identity Opinion Influence Model

5.3.1 Introduction

In the previous section, we have explored the role of the content information in opinion influence modeling. The texts as the carrier of users' opinions benefit us a lot in understanding how a person changes her/his opinions after communication. In the meanwhile, as discussed in Section 5.1, the user identity is also an important component to understand the influence process. Previous studies consider the unique user identity in a social network, and their different effects in opinion influence process are characterized by interpersonal influence. However, within the social network, apart from being a unique person, which distinguishes an individual from others (i.e., *personal identity*), a user also possesses another identity, which is her/his *social identity* [182]. The social identity specifies the extent to which individuals identify themselves in terms of group memberships. The personal identity and social identity together provide a complete user image, and affect social interactions. As Goodenough suggests: *"whenever a person interacts with another, he bases his actions on what he construes to be his own and the others' personal and social identity"* [63]. These characteristics of interpersonal and intergroup social interactions are studied in several psychological experiments [54, 158, 59]. The similar observations could also be found in the opinion influence process. For example, a user who plays as an "expert" in the discussion on a specific product would like to post the professional information and suggestions to maintain her/his position within the community. The "expert" identity would bring a person's positive influence towards other social users. It is beneficial that both users' personal identity and social identity are incorporated together to understand the opinion influence process.

Even though user dual identity would help a lot in modeling users' behaviors, their effects have not been quantitatively measured by researchers. Most of the efforts in the past were devoted to categorizing users into different social identities [112, 182]. Among these methods, network properties and exchanged textual messages are two important indicators for characterizing different social roles [2]. However, few studies tried to distinguish how different users' personalities affect their opinion behaviors. We are the first to study the effect of user dual identity in shaping the opinion influence process, and further changing people's opinion behaviors. The focus of this work is to model opinion influence by taking into account both personal identities and social identities of users.

To achieve this goal, we propose a Dual Identity based opinion Influence model (DI^2), which consists of two components. One is social role detection. The other one is opinion influence modeling. Because the criteria used to define the social identities of the users who are interested in different topics vary, it is very difficult to devise a unified standard to categorize users' social identities. We cast social identity detection to a task of user clustering. By representing users with the textual features corresponding to their opinion contents and structural features corresponding to their network properties, we divide users into different groups with different social identities. The part of opinion influence modeling follows the same way as the neural opinion influence model proposed in Section 5.2. To better depict the personalities of users, two types of influence including individual-based influence (i.e., interpersonal influence in Section 5.2) and group-based influence are both integrated in DI^2. The group-based influence enhances the individual-based influence, especially when the latter is difficult to learn from the insufficient communications between two users. Due to the interplay of social identity detection and opinion influence modeling [175], a novel joint learning framework is proposed, which has the ability to infer users' social identities and learn the dual identity based opinion influence model. Using the learned DI^2, the sentiment of a user' future opinion can then be predicted. The

[2]Social identity and social role are two interchangeable terms in this study

experiments demonstrate that compared with NIM, DI^2 has a better ability to predict users' future sentiment, especially on negative sentiment prediction.

5.3.2 Related Work in Social Identity Detection

Nowadays, it is a fact that users play different social identities (or social roles) on social media, which can be characterized as their positions, behaviors, or virtual identities. In most cases, the characteristics of social identities are pre-defined. Researchers proposed a variety of approaches to quantitatively analyze the social identities of users in different network communities. For example, [169] analyzed four types of editors in the Wikipedia community according to their edit histories and egocentric network. [80] analyzed the users of the Palins email network based on their network structure. A lot of efforts have been devoted in semi-supervised or supervised methods to detect social identity automatically. Different classifiers were developed based on textual or categorical information to predict user attributes on social media [102, 112, 117, 181, 182]. The premise of this line of research was that the social roles of users have already been labeled. However, in a real social network, the labels of peoples' social roles are usually unavailable. Also, the criteria for identifying the social roles of users change when they are applied to different social communities. This makes it difficult to annotate manually.

Another line of research focused on clustering methods, which grouped users into different clusters corresponding to different social identities. Usually, the network structure was employed to construct the feature set for clustering. The most commonly used features included the number of followers, the number of followees and etc. Besides, the equivalence between users, which measured the distance between two users [53, 20, 25, 51], homophily and triadic closure [182] were also commonly used in the existing studies. In addition to the network properties, some work characterized users according to their textual postings. Most of the research work in this direction can be seen as the extensions of the basic topic model (i.e., LDA) [19]. One notable work was Role-Author-Recipient-Topic [112], which identified a person's job title (e.g., executive assistant) in the company by analyzing the content of the emails that people sent and/or received. The model assumed that if two persons had similar probability distributions over their communication partners, they had the same managerial role. Recently, [29] proposed to combine the network and textual features together to group users according to their properties during the communication. With the similar idea, we also combine the features of network and exchanged content to characterize social identities of users in the opinion influence process.

Though the problem of social identity detection has been studied for several decades, exploration of their effects on human behavior modeling has been barely examined. In 2014, Yang et al. [175] first studied how information diffusion was influenced by the role of users. Given the detected three structural roles (i.e., opinion leaders, structural hole spanners and ordinary users) in advance, a generative model was developed to predict who would repost the message next. However, their study ignored the personal identity of users and their approach cannot be directly applied to opinion influence modeling. Different from Yang's study, we propose a novel joint learning framework to detect social roles and model opinion influence simultaneously.

5.3.3 Problem Formulation

The notations presented in Table 5.1 are also utilized in this section. It includes the network relationships F_u and the records of opinion behaviors S_u. Besides, we also utilize three profile-related features to complete the description of users in social media. They are the number of followers, the number of followees and a binary value indicating whether a user account has been verified to be authentic or not. For each user u, the network features are included in a vector $\boldsymbol{q_u}$. The problem definition becomes: given the communication records including neighboring opinion

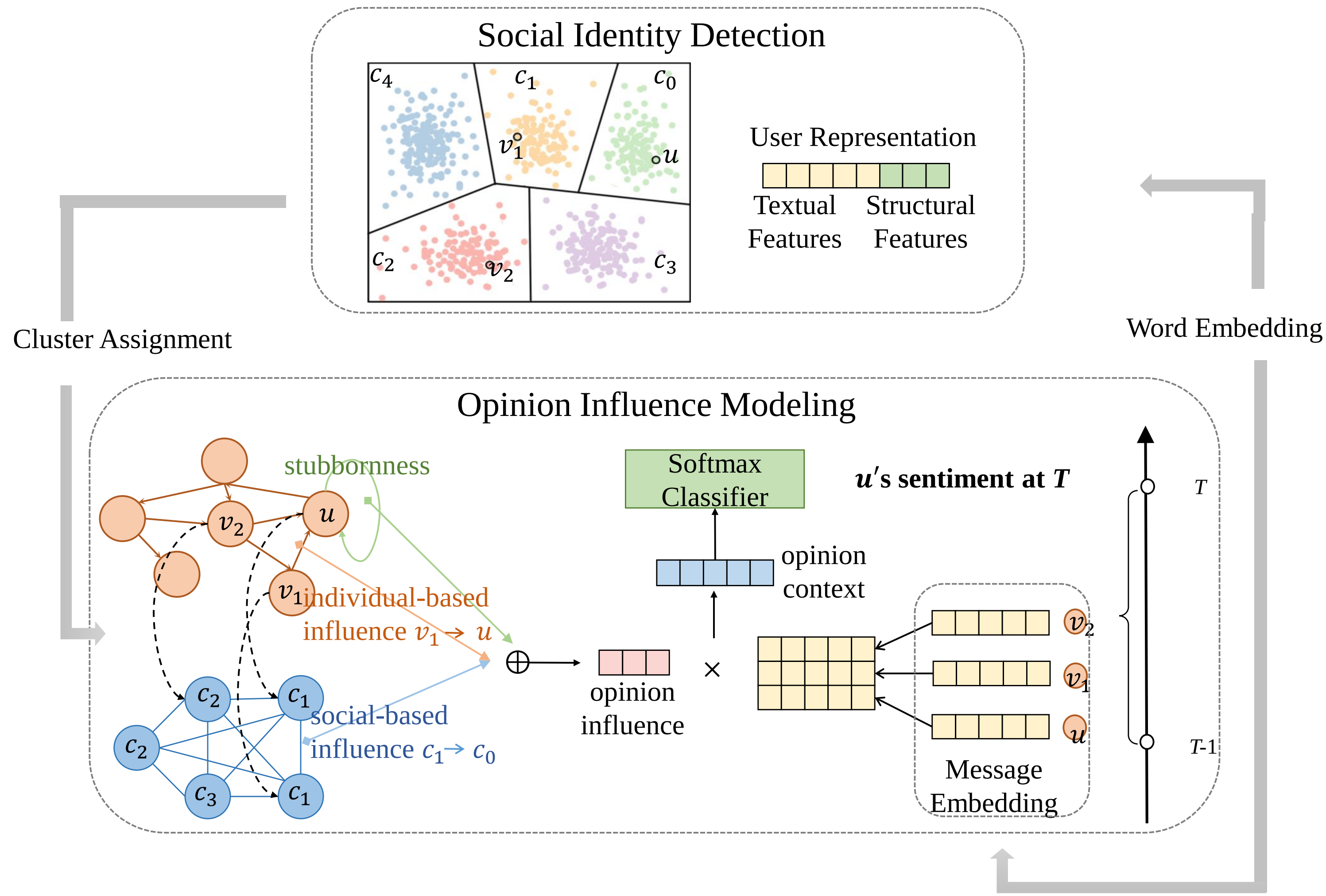


Figure 5.3: The graphical representation of DI²

information $C_u(i)$ and previous personal opinion $W_u(i-1)$, and user profiles $\boldsymbol{q_u}$, we aim to predict the future sentiment $o_u(i)$ of user u at timestamp $t_u(i)$.

5.3.4 Proposed Model

The proposed DI^2 model contains two major components, as briefly illustrated in Figure 5.3. The backbone of social identity detection is a soft K-means clustering algorithm, which aggregates users into different social identities according to their profiles. Both detected social identities combined with personal identities are considered in opinion influence process. Two types of neighboring relationships including individual-based influence and group-based influence are considered accordingly. The opinion influence model follows the same idea as proposed in Section 5.2. It is based on a neural network framework to take advantage of the powerful semantic representation ability. The posting records are powerful indicators both for user profiles and opinion changes. Thus understanding the content information is important for both tasks. We propose a joint learning framework where social identity detection and opinion influence modeling are learned simultaneously by sharing the same word embeddings.

Social Identity Detection

User Representation Usually, the features of users are reflected in two aspects, i.e., the textual features, which describe a user's personal interests and opinions, and the structural features, which reflect their structural positions within the whole social network [29]. We capture an individual's personal interests from her/his past posting records by averaging all the messages u posted in the sequence $S_u = \{W_u(1), \cdots, W_u(n_u)\}$. The word embedding technique, which is used in the previous section to avoid the curse of dimensionality problem, is also used here to represent users' textual features. The textual feature vector $\boldsymbol{h_u}$ is represented as:

$$\boldsymbol{h_u} = \frac{\sum_{i=1}^{n_u}(\sum_{j=1}^{|W_u(i)|}\boldsymbol{\Phi} W_{u,j}(i-1))}{n_u} \tag{5.10}$$

The structural features, which depict the network status of an individual are included in the vector $\boldsymbol{q_u}$. Given the textual features and structural features, we represent each user with a d_u-dimensional vector $\boldsymbol{x_u}$.

$$\boldsymbol{x_u} = [\boldsymbol{h_u}; \boldsymbol{q_u}] \tag{5.11}$$

User Clustering The popular K-means clustering algorithm [29] is utilized to aggregate users with similar characteristics into the same group. The center of cluster c_j is denoted by the vector $\boldsymbol{\theta_j}$, and its dimension is d_u, which is same as the dimension of user representation. The total number of social identities is K. Intuitively, a user may play multiple social identities with respect to different communities or groups. Here, we use the soft version of K-means [17, 50] to allow each user to be assigned to multiple clusters with a probabilistic distribution $\boldsymbol{z_u}$ over all the clusters. The probabilities are calculated by the distances between a user and K cluster centers, i.e.,

$$\boldsymbol{z_u} = softmax(-\boldsymbol{l_u}) \tag{5.12}$$

where each element of $\boldsymbol{l_u}$ is represented by $l_{uj} = \|\boldsymbol{x_u} - \boldsymbol{\theta_j}\|_2$

Note that to better illustrate the proposed model in Figure 5.3, we only show one assigned cluster for each user instead of presenting her/his probability distribution.

Dual-identity Opinion Influence Model

In Section 5.2, we characterize opinion influence as interpersonal influence between each user pair. The interpersonal influence modeling is based on users' personal identities. Though personal identity is an important factor during social interactions, the investigation of personal identity alone is not necessarily complete to construct an individual's personal image. In DI^2, we explore both personal identity and social identity in opinion influence modeling to complete the user portrait during social interactions. Considering the dual identity, the neighboring relationship can be characterized into two types of influence, i.e., individual-based influence and group-based influence. The individual-based influence corresponds to personal identity and the group-based influence corresponds to social identity.

The individual-based influence carries the same meaning as used in the previous studies, which depicts the part of influence occurring when each person possesses its own personal identity. It is represented by the matrix $\boldsymbol{A} \in \mathbb{R}^{N \times N}$, and constrained by the network structure. Additionally, the group-based influence is denoted by a matrix $\boldsymbol{B} = [\beta_{ij}] \in \mathbb{R}^{K \times K}$, where the influence from group c_i on group c_j is β_{ij}. As a user may play multiple social identities under a probabilistic distribution $\boldsymbol{z_v}$, the group-based influence between two users v and u is:

$$e_{uv} = \boldsymbol{z_u} \boldsymbol{B} \boldsymbol{z_v}^T \tag{5.13}$$

Both individual-based influence and group-based influence reflect the neighboring relationship between each pair of users, and the integrated influence is represented in Equation 5.14.

$$inf_{uv} = sigmoid(\lambda_{uv}) * \alpha_{uv} + (1 - sigmoid(\lambda_{uv})) * e_{uv} \tag{5.14}$$

$sigmoid(\lambda_{uv})$ denotes the probability to balance the contribution of individual-based influence and group-based influence. The *sigmoid* function ensures the probability between 0 and 1. On social media, the interactions between different user pairs are various. For example, the group-based influence dominates when two users know few about each other and have few interactions. In contrast, when two users frequently communicate with each other, they tend to be very familiar with the other's personalities, and the individual-based influence would hold a dominant position. Thus, we set $\lambda_{uv} = \lambda_0 + \omega * f_{uv}$, where λ_0 is a parameter shared by all users, f_{uv} is the interaction frequency between each user pair, and ω denotes the weight. The interaction frequency f_{uv} is measured by the percentage of the messages that u changes opinions after receiving the opinions from v.

In addition to neighboring opinion influence, a user also tends to insist on her/his prior opinion during the communication [44]. We use α_{u0} to denote the degree of u's stubbornness on the prior opinion. Similar to the neural opinion influence model introduced in Section 5.2, an opinion context vector is constructed by taking the prior messages from a user's neighbors and her/his own as the input, and integrating the messages with neighboring opinion influence and personal stubbornness.

$$\boldsymbol{c_u}(i) = \alpha_{u0}\boldsymbol{p_u}(i-1) + \sum_{v} inf_{uv}\boldsymbol{s_u^v}(i) \tag{5.15}$$

where the personal prior opinion $\boldsymbol{p_u}(i-1)$ and the neighboring opinion $\boldsymbol{s_u^v}$ are defined in Equation 5.2 and 5.3 accordingly.

Given the opinion context vector $\boldsymbol{c_u}(i)$, the output layer is a softmax function to output the probabilities over all types of sentiment.

$$P(o_u(i)|\boldsymbol{c_u}(i)) = softmax(\boldsymbol{V}\boldsymbol{c_u}(i) + \boldsymbol{b}) \tag{5.16}$$

where $\boldsymbol{V} \in \mathbb{R}^{O\times d_w}$, and $\boldsymbol{b} \in \mathbb{R}^{O}$. O is the number of sentiment polarities, which is 3.

Joint Learning Framework

The loss function of DI^2 contains two terms. They are the loss function of neural sentiment classification and the loss function for soft K-means clustering.

$$\epsilon = -\sum_{u=1}^{N}\sum_{i=1}^{n_u} \log P(o_u^*(i)) + \sum_{u=1}^{N}\sum_{j=1}^{K} z_{uj}\|\boldsymbol{x_u} - \boldsymbol{\theta_j}\|_2 \tag{5.17}$$

where $o_u^*(i)$ represents the sentiment label of u's message at $t_u(i)$.

Algorithm 5.1 Joint learning algorithm

Initialization: Maximum training iteration R, word representation $\mathbf{\Phi}$, cluster center $\mathbf{\Theta}$, former cluster center $\mathbf{\Theta'}$, individual-based influence $\boldsymbol{A}$, group-based influence $\boldsymbol{B}$, balance weight λ_0, weight of interaction frequency ω, output parameters $\boldsymbol{V}$, $\boldsymbol{b}$ and smooth weight δ

for iteration i =1,2,...,R **do**

  for each batch bs **do**

    Given the updated $\mathbf{\Phi}$ in previous batch, compute user representation $\boldsymbol{x_u}$ for $u \in BU$, where BU is the set of users included in the batch bs

    Calculate the soft assignment $\boldsymbol{z_u}$ for each user according to Equation 3.

    Compute the new cluster center $\hat{\boldsymbol{\theta}}_{\boldsymbol{j}} = \frac{1}{|BU|} \sum_{u \in BU} z_{uj} \boldsymbol{x_u}$

    Smooth cluster centers $\mathbf{\Theta} = \delta \hat{\mathbf{\Theta}} + (1 - \delta) \mathbf{\Theta'}$

    set $\mathbf{\Theta}$ as $\mathbf{\Theta'}$

  end for

  Update the parameters $\mathbf{\Phi}$, $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{V}$ and $\boldsymbol{b}$, λ_0, ω with the SGD algorithm given the cluster center $\mathbf{\Theta}$

end for

The learning algorithm is summarized in Algorithm 5.1. Mini-batch is used during model training. In the forward pass, given the newly updated word embedding matrix $\mathbf{\Phi}$ in the previous batch, we update the user representations and their probabilistic distributions over all the clusters in the current mini-batch. When the cluster assignments of users have been changed, the cluster centers $\hat{\mathbf{\Theta}}$ should be changed accordingly. Considering $\hat{\mathbf{\Theta}}$ is computed for the users within the current batch, it cannot represent the partition over all users. To address the issue, we use the online updating algorithm [107] to update cluster centers, i.e., the updated cluster center $\mathbf{\Theta}$ is the combination of the former cluster center $\mathbf{\Theta'}$ and the cluster center $\hat{\mathbf{\Theta}}$ learned from the current batch. In the backward pass, given the estimation of cluster centers $\mathbf{\Theta}$, we update word embeddings and the other parameters in the opinion influence model with the stochastic gradient descent (SGD) algorithm. The training phase stops when the training error has a decrease less than 1 or reaches the maximum iteration $R = 100$.

5.3.5 Experiments and Discussions

Experimental Set-up

In Section 5.2, the superior performance of the neural opinion influence model to the state-based opinion influence models has been demonstrated. In this section, we focus on investigating the effectiveness of user dual identity in opinion influence, and compare DI^2 with another two content-based opinion influence models, which individually consider personal identity and social identity. Besides, we also compare DI^2 with the pipeline framework to demonstrate the effectiveness of the proposed joint learning framework.

PI^2: It is a variation of DI^2 involving personal identity alone. It is the neural opinion influence model (NIM) proposed in Section 5.2, and has achieved better performance than the state-of-the-art opinion influence models.

SI^2: It is the other variation of DI^2, which considers social identity alone. In this setting, the opinion influence inf_{uv} between two users is only decided by the group-based influence e_{uv} identified in Equation 5.13. Thus, the opinion context vector used in the output layer Equation 5.16 is defined as:

$$\boldsymbol{c_u}(i) = e_{uu}\boldsymbol{p_u}(i-1) + \sum_{v} e_{uv}\boldsymbol{s_u^v}(i) \tag{5.18}$$

PIPE_DI^2: Like DI^2, PIPE_DI^2 bears the ability to integrate user dual identity into opinion influence modeling. Different from DI^2, which clusters users' social identities and captures their opinion influence in a unified jointly learning framework, PIPE_DI^2 employs a pipeline framework [175]. It divides the method into two independent parts. The first step of the framework is to use the soft K-means method to aggregate users into groups. In this step, each user is represented by the network features and the bag-of-words textual features. Given the cluster assignments of users detected from the first step, the second step is to employ the opinion influence model to learn both individual-based influence and group-based influence for sentiment prediction.

For DI^2, we set the dimension of the word embedding d_w as 30, the initial values of the balance weights λ_0 and ω as 0 and 1, and the smooth weight δ as 0.5. We experimentally set the cluster number K to 5. It is also the same as the number of the influence roles proposed in [29] by analyzing users' characteristics in communication. To make a fair comparison, we use the same parameter setting for PI^2, SI^2, PIPE_DI^2 and DI^2.

Performance Evaluation

As reported in Table 5.3, DI^2 performs the best in almost all evaluation metrics. Based on the results, some important findings are concluded as follows.

First, we find that dual identity better represents users. The better performance of DI^2 compared against PI^2 and SI^2 demonstrates that dual identities provide a more completed representation for each user, and the two types of opinion influence derived from dual identities better capture the relationships between user pairs during the communication. Meanwhile, PI^2 outperforms SI^2, which indicates that personalities of users play a dominant role during the process of opinion formation.

Second, we observe that the joint learning framework brings benefits. DI^2 with the joint learning framework is superior to PIPE_DI^2 with the pipeline framework. With the novel joint learning framework, the tasks of opinion influence modeling and social identity detection benefit each other by learning interactively.

Finally, DI^2 copes better with the negative sentiment. Compared with positive and neutral sentiment prediction, negative sentiment prediction achieves the lowest

Table 5.3: Performances on three products.

(a) Samsung Galaxy

Method	Accuracy	F1_Neg	F1_Neu	F1_Pos
PI^2	0.6590	0.2075	**0.7306**	0.6357
SI^2	0.6196	0.0560	0.7138	0.5280
$PIPE\_DI^2$	0.6542	0.2452	0.7214	0.6252
DI^2	**0.6660**	**0.2874**	**0.7304**	**0.6450**

(b) Xbox

Method	Accuracy	F1_Neg	F1_Neu	F1_Pos
PI^2	0.5694	0.2346	0.6272	0.6002
SI^2	0.5497	0.0670	0.6135	0.4973
$PIPE\_DI^2$	0.5447	0.2698	**0.5954**	0.5812
DI^2	**0.5823**	**0.3017**	**0.5962**	**0.6137**

(c) PlayStation

Method	Accuracy	F1_Neg	F1_Neu	F1_Pos
PI^2	0.6653	0.1301	0.7813	**0.5136**
SI^2	0.6380	0.0752	0.7659	0.3474
$PIPE\_DI^2$	0.6477	0.1586	0.7686	0.4737
DI^2	**0.6751**	**0.1957**	**0.7904**	0.5028

results. Compared to the second best results of the content-based models, the F-measures of DI^2 on negative sentiment prediction are greatly improved by 17.2%, 11.82% and 23.39% on the topics "Samsung Galaxy", "Xbox" and "PlayStation", respectively. As shown in Table 4.2, less than 20% of tweets express the negative sentiment. The combination of dual identities allows DI^2 to better "understand" negative opinion formation when the communication records are insufficient.

5.4 Chapter Summary

In this chapter, we focus on the exploration of content information in opinion influence modeling. The significant improvements of the proposed model compared with existing state-based models demonstrate the powerful ability of content information in understanding the opinion influence process. Furthermore, we extend the neural opinion influence model by capturing the complete user portfolios during their interactions. It integrates together user personal identity and social identity to model the opinion influence process. Considering the interplay of social identity detection and opinion influence modeling, a joint learning framework is proposed, which creatively clusters users into different groups, as well as learns neighboring relationships between them. The effectiveness of the joint learning framework is demonstrated in the experiments.

Chapter 6

Content-based Sequential Opinion Influence Modeling

6.1 Chapter Overview

In the previous two chapters, we explore the role of temporal properties and the content information in opinion influence modeling separately and demonstrate their effectivenesses by conducting comprehensive experiments. It is natural to go further a step to understand the opinion influence process by collectively exploring all of them. In this chapter, we present a unified framework built to deliver a comprehensive understanding for the opinion influence process during the communication.

To this end, the proposed model should have the ability to track the content-based opinion dynamics and uncover the correlation of users from their temporal interactions. The content-based behaviors have been effectively processed by the word embedding techniques in Chapter 5. In this way, the opinion is described as a continuous vector. To track the temporal opinion behaviors, a temporal influence model is proposed in Chapter 4. The model based on the coupled Markov chain is able to cope with the sequence of discrete opinion states. However, it cannot handle the opinion states in the form of continuous vectors. In addition, the temporal influence model utilizes several predefined assumptions to model temporal properties, but lacks the ability to uncover temporal dependencies in opinion behaviors.

The Recurrent Neural Network (RNN) model is another effective architecture for sequence modeling. It does not make any assumption about the sequence, but learns the sequential properties from the data. The superiority of the recurrent neural network over the Markov chain also has its root in the greater representation power of neural networks, which provides the possibility to combine the content information into sequence modeling. Extended from the basic RNN architecture, the Gated RNN [36] and the Long Short Term Memory network (LSTM) have been developed with the ability to select when to drop or save the historical information. The idea of the reset gate and the update gate is also suitable for modeling the decision process of human beings when they form opinions. For instance, the users in social media tend to forget the irrelevant histories.

Taking advantage of the recurrent neural network, we propose a framework, named Content-based Sequential Opinion Influence framework (CSIM) to model the sequential opinion influence process. The opinion dynamics of each user is recorded with an individual RNN chain. The internal state of a RNN contains all the information related to one's future sentiment. The opinion state transition is decided by both long-term communication records and recent neighboring opinions, and is implemented in the recurrent unit. Thus, the neural opinion influence model proposed in Chapter 5 is extended to a setting where the opinion influence process is sequentially captured from the opinion dynamics.

Additionally, we explore different architectural designs of the output structure depending on different prediction strategies, and the alternative models CSIM_S and CSIM_W are developed. In the previously proposed opinion influence models, the sentiment of opinion, as the target of prediction is directly utilized as the labels during the training process. In this chapter, the same strategy is still adopted in the model CSIM_S. Though this strategy provides a fast and straightforward way to correlate the related information to the future sentiment, the parameter learning process may be harmed by the summarized sentiment labels. It is because labels cannot sufficiently represent the opinion information included in the messages. We furthermore deliberate the output structure, and propose a new prediction strategy in the model CSIM_W. Instead of predicting the target of sentiment, the model first predicts the intermediate results (i.e., opinion words) and then obtains the sentiment by summing the sentiment scores of all predicted opinion words. The learning of CSIM_W is strictly constrained by the fine-grained opinion words, which can effectively avoid the negative influence of summarized labels.

To demonstrate the effectiveness of the proposed model, we compare it with the models introduced in previous chapters, including temporal value-based influence models and content-based neural influence models. The models with both prediction strategies perform better than the compared models. Especially, CSIM_W achieves the best performance, which demonstrates the effectiveness of the proposed opinion word-based prediction technique. We also present a number of experiments to analyze the opinion influence problem from a variety of viewpoints.

6.2 Related Work in Recurrent Neural Network

The recurrent neural network (RNN) was first proposed by [52] to solve the problem of modeling sequences with arbitrary lengths. Compared with the classical feed-forward neural network (FNN), it incorporates the extra neurons and connects them to a hidden layer like the other input neurons. The extra neurons are context neurons that record histories of sequences. The output of FNN is fed into the context unit, which is in turn sent to the hidden layer in the next iteration. Because of the advantages of RNN in recording historical memories, it has been widely applied in a variety of applications since proposed.

When the length of sequence increases, RNN suffers from the vanishing or exploding gradient problem. To alleviate the problem, a lot of variants of RNN including LSTM [76] and GRU [36] are proposed. LSTM replaces the hidden unit with a memory block, which contains a memory cell and three gates. The three gates including the input, output and forget gates provide the write, read and reset operations for the cells. The hidden units allow the memory cells to access and forget the historical information. However, the architecture of LSTM is complex and improves the cost of training. In this situation, GRU [36] is proposed to balance the effectiveness and efficiency in sequence modeling by maintaining two gates, i.e., reset and update gates. Both RNN and its variants have been applied in a variety of applications, including machine translation [36, 9, 172], dialogue generation [131, 147] and opinion mining [153, 108]. Among these applications, opinion mining [92, 155, 153] is the one most related to our work. For example, Tang et al. [153] proposed a gated recurrent neural network to classify the sentiment of a document. The RNN was used to compress the document to a single vector, which was further utilized for classifying the sentiment. However, the opinion influence mechanism behind the sentiment formation has not been explored. To the best of our knowledge, our work is the first to apply the RNN framework to capture the sequential properties of opinion dynamics.

6.3 Problem Formulation

The notations used in this chapter are presented in Table 6.1, including the network structure and posting records. With the above notations, the opinion prediction problem is defined as: given the communication histories before timestamp $t_u(i-1)$, including personal posting records $S_u(< i)$ and neighboring posting records $C_u(< i)$, our objective is to predict the future sentiment $o_u(i)$ of user u at next timestamp $t_u(i)$. The optimization problem becomes to maximize the following objective function

Table 6.1: Definition of notations

Notation	**Description**
V	a set of $\|V\| = N$ users.
F_u	a set of u's neighbors in the network. The size is m_u.
$W_u(i)$	opinion words included in u's i-th message
$o_u(i)$	the sentiment of u's i-th message.
$t_u(i)$	the time u publishes the i-th message.
$S_u(< i) \triangleq \{W_u(1), \cdots, W_u(i-1)\}$	a set of messages posted by u before $t_u(i)$
$C_u(i) = \{C_u^1(i), \cdots, C_u^v(i), \cdots, C_u^{m_u}(i)\}$	a set containing all the messages posted by u's neighbors between $t_u(i-1)$ and $t_u(i)$
$C_u(< i) \triangleq \{C_u(1), \cdots, C_u(i-1)\}$	messages posted by neighbors before $t_u(i)$

consisting of the estimation of all user opinion behaviors.

$$\mathcal{L}(S) = \sum_{u=1}^{N} \sum_{i=1}^{n_u} \log P(o_u(i)|C_u(< i), S_u(< i))$$

6.4 Proposed Framework

6.4.1 Motivation

Due to the sequential properties of the opinion influence process, we propose to develop a sequential model to temporally track the content-based users' behaviors and capture their opinion influence. At each time stamp, three crucial factors should be considered for shaping one's future opinion, i.e., 1) personal prior opinion influence, 2) new coming neighboring opinion influence, and 3) historical communication influence. The first two factors have been studied in Section 5.2 and are captured by transforming the opinion words into the word embeddings. The word embedding techniques have been proved to be successful in not only avoiding the "curse of dimensionality" problem, but also uncovering the opinion influence process from the semantic information. Then, the next challenge is how to represent the historical communication records, and integrate it into opinion influence modeling.

During the continuous communication over the social network, all past exchanged information can produce effects on users when they update future opinions. The historical opinion influence at $t_u(i)$ can be represented as a function of all the neighboring information that u receives before $t_u(i)$ and all u's personal prior opinions.

$$\boldsymbol{l_u}(i) = f(C_u(< i), S_u(< i)) \tag{6.1}$$

We aim to find a function f having the ability to integrate the past information iteratively over time. The RNN framework provides the possibility to maintain a vector for each time stamp by integrating the information included in historical sequences. Thus, it motivates us to employ the RNN architecture for capturing influence from historical communication. We develop a Content-based Sequential Opinion Influence Model (CSIM) that employs the RNN techniques to encode the historical communication, and integrates it with personal prior opinion and the new coming neighboring opinions for discovering sequential opinion influence.

In addition, based on the CSIM framework, we propose two alternative models with different prediction strategies. The first one is the CSIM_S model using the sentiment based prediction strategy. CSIM_S obtains the probability distribution over the three sentiments directly from the internal opinion state. In this setting, the summarized sentiment category labels are taken as the prediction targets. The model learns the influence parameters by exploring the correlation between the neighboring opinion information a user receives and her/his future sentiment. The summarized sentiments only reflect partial opinion information included in one's future message. The learning process constrained by the summarized sentiment labels may not accurately estimate the influence parameters, which may lead to lower accuracies of prediction results. To improve the accuracy, we further develop the CSIM_W model using the opinion word based prediction strategy. CSIM_W takes the fine-grained opinion words as the prediction targets. The learning of the model is strictly constrained by utilizing the exact opinion words as supervised labels instead of the pre-summarized sentiments. The learned parameters by CSIM_W can better describe the mechanisms underlying opinion formation than the ones learned by CSIM_S. The overall sentiment can then be obtained by summing together the sentiment values of all the predicted opinion words.

Sequential Modeling of Opinion Dynamics

As mentioned in the previous chapter, we obtain the contextual vector $\boldsymbol{c_u}(i)$ at each time stamp by combining the vectors of one's personal prior opinion $\boldsymbol{p_u}(i-1)$ and her/his neighboring opinion vector $\boldsymbol{s_u}(i)$. The contextual vector summarizes the most

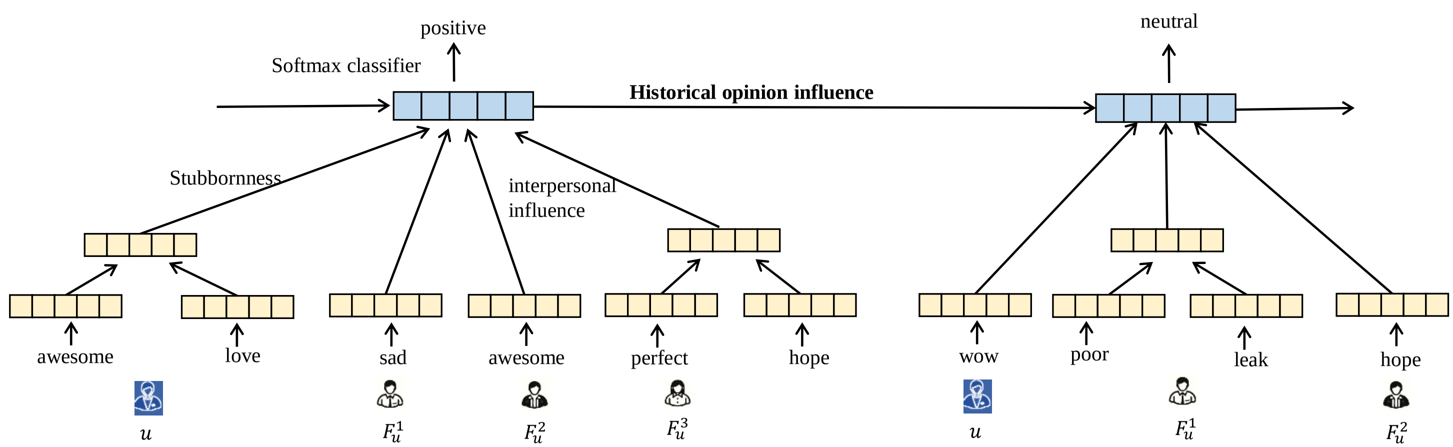


Figure 6.1: The graphical representation of CSIM_S. It predicts the sentiment category labels directly in the output layer.

recent information a user would like to take as the reference for her/his next opinion. Thus, the current internal opinion state $\boldsymbol{h_u}(i)$ could be computed by combining $\boldsymbol{c_u}(i)$ with the information already processed, i.e., the historical opinion influence vector $\boldsymbol{l_u}(i-1)$.

$$\boldsymbol{h_u}(i) = f(\boldsymbol{l_u}(i-1), \boldsymbol{c_u}(i)) \tag{6.2}$$

where $\boldsymbol{h_u}(i) \in \mathbb{R}^{d_h}$, and d_h is the dimension of the internal state. The current internal opinion state $\boldsymbol{h_u}(i)$ compresses all the relevant information during the communication, which can be taken as historical opinion influence $\boldsymbol{l_u}(i)$ for updating the next opinion at $t_u(i+1)$. Hence, $\boldsymbol{l_u}(i-1)$ could be replaced by the internal state $\boldsymbol{h_u}(i-1)$ at $t_u(i-1)$, and the updating rule becomes:

$$\boldsymbol{h_u}(i) = f(\boldsymbol{h_u}(i-1), \boldsymbol{c_u}(i)) \tag{6.3}$$

The first internal opinion state $\boldsymbol{h_u}(0)$ is initiated with a zero vector.

The transformation function f for a standard recurrent neural network is usually a non-linear function, such as hyperbolic tangent:

$$\boldsymbol{h_u}(i) = tanh(\boldsymbol{W_r}[\boldsymbol{h_u}(i-1); \boldsymbol{c_u}(i)] + \boldsymbol{b_r}) \tag{6.4}$$

where $\boldsymbol{W_r} \in \mathbb{R}^{d_h \times (d_h + d_w)}$, and $\boldsymbol{b_r} \in \mathbb{R}^{d_h}$, and d_w is the dimension of the opinion word vector.

Though the standard recurrent neural network could successfully handle the history information, it suffers from the vanishing gradient problem [15, 135]. The problem is that, in some cases, the gradient will be vanishingly small, and prevent the parameters to change. In the worst case, this may completely stop the neural network from further training. To address this problem, several complex transformation functions or recurrent units are proposed, such as the Long Short-Term Memory (LSTM) [76] and the Gated Recurrent Unit (GRU) [36]. These recurrent units effectively allow the internal state to drop the irrelevant information for the future state, or to reserve the relevant history, thus, allowing a more compact representation. In our proposed model, we choose to use GRU, which has been proved more effective than the simple transformation function used in RNN and has an affordable computation cost [39]. The GRU is formally expressed as follows:

$$\boldsymbol{r_u}(i) = sigmoid(\boldsymbol{I_r c_u}(i) + \boldsymbol{H_r h_u}(i-1)) \tag{6.5}$$

$$\boldsymbol{d_u}(i) = sigmoid(\boldsymbol{I_d c_u}(i) + \boldsymbol{H_d h_u}(i-1)) \tag{6.6}$$

$$\boldsymbol{g_u}(i) = tanh(\boldsymbol{I_c c_u}(i) + \boldsymbol{H_c}(\boldsymbol{r_u}(i) \cdot \boldsymbol{h_u}(i-1))) \tag{6.7}$$

$$\boldsymbol{h_u}(i) = (1 - \boldsymbol{d_u}(i)) \cdot \boldsymbol{h_u}(i-1) + \boldsymbol{d_u}(i) \cdot \boldsymbol{g_u}(i) \tag{6.8}$$

where $\boldsymbol{I_r}, \boldsymbol{I_d}, \boldsymbol{I_c} \in \mathbb{R}^{d_h \times d_w}$, and $\boldsymbol{H_r}, \boldsymbol{H_d}, \boldsymbol{H_c} \in \mathbb{R}^{d_h}$.

Two important gates in GRU control the long-term and short-term memories, respectively. The reset gate $\boldsymbol{r_u}$ is responsible for dropping the part of information that is not relevant to the future. It is similar to the forgetting scheme of human beings. As time goes, it is difficult for people to remember all past events, and they will inevitably forget some old stories that are irrelevant to their future decisions. Meanwhile, the update gate $\boldsymbol{d_u}$ is to control whether or not the new coming information should be kept in the internal state $\boldsymbol{h_u}(i)$. When the short-term received information is useless for a user's future opinions, the update gate will be pushed close to 0, and the previous internal state $\boldsymbol{h_u}(i-1)$ will be kept as the current internal state.

6.4.2 Prediction Strategy

Based on the CSIM framework, we now introduce the models utilizing two alternative prediction strategies, i.e., the sentiment based prediction strategy (CSIM_S) and the opinion word based prediction strategy (CSIM_W).

Sentiment based Prediction

Output Structure: Given the internal opinion state $\boldsymbol{h_u}(i)$, which represents the features for predicting future sentiment, the output layer for the CSIM_S is a softmax function σ that output probabilities over all categories of sentiment. The graphical representation of CSIM_S is presented in Figure 6.1.

$$P(o_u(i)|\boldsymbol{h_u}(i)) = \sigma(\boldsymbol{V}\boldsymbol{h_u}(i) + \boldsymbol{b})) \tag{6.9}$$

and

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \tag{6.10}$$

where $\boldsymbol{V} \in \mathbb{R}^{K \times d_h}$, and $\boldsymbol{b} \in \mathbb{R}^{K}$. K is the number of sentiment polarities, which is set to 3.

Learning: The CSIM_S is parameterized by the social influence factors $\boldsymbol{\alpha}$, the word representation $\boldsymbol{\Phi}$ for each opinion word, the parameters for gated recurrent unit $\boldsymbol{I_r}$, $\boldsymbol{I_d}$, $\boldsymbol{I_c}$, $\boldsymbol{H_r}$, $\boldsymbol{H_d}$, $\boldsymbol{H_c}$, and the output parameters $\boldsymbol{V}$ and $\boldsymbol{b}$. The objective function to be maximized is the log-likelihood of the sequences of opinion sentiments. It is defined as below.

$$\mathcal{L}(S) = \sum_{u=1}^{N} \sum_{i=1}^{n_u} \log P(o_u(i)|C_u(< i), S_u(< i)) \tag{6.11}$$

We compute the gradients using the back-propagation through time (BPTT) algorithm [145]. Then the parameters are updated by Adam [94], i.e., a variant of the stochastic gradient descent (SGD) algorithm.

Opinion Word based Prediction

The model CSIM_S takes the sentiment category label $o_u(i)$ as the ground-truth in learning (see Equation 6.11). The label only represents the sentiment polarity

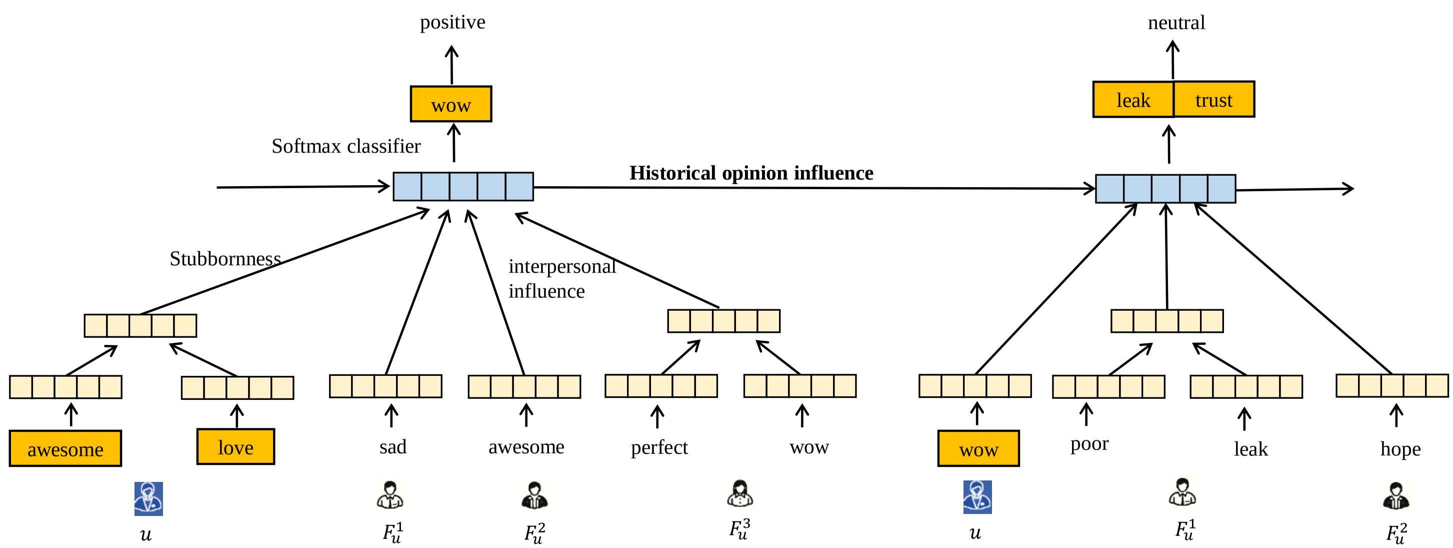


Figure 6.2: The graphical representation of CSIM_W. It predicts the opinion words in the output layer first and then obtains the sentiment labels.

of an opinion expression. It cannot reflect the opinion content information. Such summarized sentiment may harm the understanding of opinion dynamics and thus affect the accuracy of the predicted results. For example, there are two users A and B with the relationship that A follows B. One day, A posts a message:

"I really *like* the design of new Samsung Galaxy Note".

After reading it, B posts:

"I *like* the new Samsung Galaxy. *Sadly*, no store here to buy it."

Summarized by the lexicon-based sentiment analysis method, the sentiment of B's posting is negative. Because of the positive sentiment of A's posting, the learned influence between A and B is negative. However, the fact is that the likeness for the new product is transmitted from user A to user B. Thus the influence between two users should be positive. Another prediction strategy is the opinion word prediction strategy, which is integrated in the model CSIM_W. A major benefit of this particular design is that after training with the opinion words, the inaccuracies introduced by the sentiment labels can be reduced. Besides, the prediction of opinion words also provides more detailed information about the future trending views about a topic. The graphical representation of CSIM_W is presented in Figure 6.2.

Output Structure: CSIM_W takes the internal opinion state $\boldsymbol{h_u}(i)$ to predict all opinion words within the tweet instead of the sentiment. More generally, we formulate the problem as predicting a probability distribution over all opinion words, and those with the high probabilities are taken as the resulting opinion words in the next tweet. In this setting, the output layer could be presented as follows:

$$\boldsymbol{z_u}(i) = P(\boldsymbol{w_u}(i)|\boldsymbol{h_u}(i)) = softmax(\boldsymbol{V_w}\boldsymbol{h_u}(i) + \boldsymbol{b_w}) \tag{6.12}$$

where $\boldsymbol{V_w} \in \mathbb{R}^{d_k \times d_h}$, and $\boldsymbol{b_w} \in \mathbb{R}^{d_k}$. d_k is the total number of all opinion words occurring in the opinion behaviors for all users. $\boldsymbol{z_u}(i)$ is the predicted probability distribution over all opinion words.

Learning with Fine-grained Supervision: We first formulate the probability distribution of the target opinion word set $W_u(i)$ as $\boldsymbol{q_u}(i)$. $\boldsymbol{q_u}(i) \in \mathbb{R}^{d_k}$ and it satisfies that for $j \in W_u(i), q_u(i)(j) = 1/|W_u(i)|$, and for $j \notin W_u(i), q_u(i)(j) = 0$.

In this setting, we use the generalization of multinomial logistic loss as the objective function to minimize the Kullback-Leibler divergence [41] between the predicted distribution $\boldsymbol{z_u}(i)$ and the target distribution $\boldsymbol{q_u}(i)$:

$$\mathcal{L} = \sum_{u=1}^{N} \sum_{i=1}^{n_u} \mathbf{KL}(\boldsymbol{q_u}(i) || \boldsymbol{z_u}(i)) \tag{6.13}$$

where

$$\mathbf{KL}(\boldsymbol{q_u}(i) || \boldsymbol{z_u}(i)) = \sum_j q_u(i)(j) log \frac{q_u(i)(j)}{z_u(i)(j)}$$

The parameters including $\boldsymbol{\alpha}, \boldsymbol{\Phi}, \boldsymbol{I_r}, \boldsymbol{I_d}, \boldsymbol{I_c}, \boldsymbol{H_r}, \boldsymbol{H_d}, \boldsymbol{H_c}$ are the same parameters used in CSIM_S, and the parameters of the output layer $\boldsymbol{V_w}$ and $\boldsymbol{b_w}$ are specific for CSIM_W. Similar to the learning process of CSIM_S, the gradients are computed by the BPTT algorithm, and all the parameters are updated with the Adam algorithm.

6.5 Experiments and Discussions

6.5.1 Compared Methods

To comprehensively compare the effectiveness of the proposed model and explore how different components affect opinion influence modeling, we provide the results of all the compared methods used in the previous section, as well as the proposed models in Chapter 4 and 5.

We summarize the compared methods as follows.

1) Value-based methods, which include the Degroot model, the Flocking model, the AsLM model and the Voter model. The details of the models are described in Section 4.5.2.

2) Value-based temporal model is the temporal opinion influence model (TIM) proposed in Chapter 4. It only considers the temporal properties of the opinion influence process, but ignores the content information included in the exchanged messages.

3) Content-based model is the neural opinion influence model (NIM) proposed in Chapter 5, which employs the content information to model opinion influence. However, the model only considers the most recently received information without the exploration of the temporal properties of the opinion influence process. Based on the NIM framework, we develop two variants according to different prediction strategies proposed in this chapter. NIM_S uses the sentiment based prediction strategy in the output layer and NIM_W uses the opinion word based prediction in the output layer.

For the content-based sequential models CSIM_S, CSIM_W, NIM_S and NIM_W, we set the dimension d_w of word representation to 30. In the proposed models CSIM_S and CSIM_W, the dimension of the internal opinion state d_h is 30. The other parameters of each model are set for their best performances experimentally.

Similar to the previous two chapters, the performance is still demonstrated on the model's ability of sentiment prediction and evaluated in terms of the prediction accuracy and the F-score over three sentiment types.

6.5.2 Experimental Results

Overall Performances

The experimental results on sentiment prediction are reported in Table 6.2. CSIM_W and CSIM_S almost outperform the other compared methods in terms of all evaluation metrics on three topics. The experimental results demonstrate that the content-based representation of opinion and the integration of historical communication in CSIM are able to better describe how an individual is influenced by their neighbors

Table 6.2: Performances on three products.

(a) Samsung Galaxy

Method		Accuracy	F1_Neg	F1_Neu	F1_Pos
Value-based Model	Degroot	0.5772	0.1688	0.6857	0.4408
	Flocking	0.5819	0.2414	0.6939	0.3950
	AsLM	0.5481	0.1555	0.6826	0.4597
	Voter	0.5822	0.1773	0.6593	0.5425
Value-based Sequential Model	TIM	0.6576	0.1683	0.7276	0.6305
Content-based Model	NIM_S	0.6590	0.2075	0.7306	0.6357
	NIM_W	0.6928	0.2179	0.7878	0.6269
Content-based Sequential Model	CSIM_S	0.6662	0.1981	0.7389	0.6359
	CSIM_W	**0.7093**	**0.2534**	**0.7976**	**0.6620**

Table 6.2: Performances on different products.

(b) Xbox

Method		Accuracy	F1_Neg	F1_Neu	F1_Pos
Value-based Model	Degroot	0.4893	0.1801	0.5897	0.4265
	Flocking	0.4407	0.0861	0.0861	0.2298
	AsLM	0.4935	0.1963	0.7233	0.3056
	Voter	0.5116	0.1861	0.5471	0.5308
Value-based Sequential Model	TIM	0.5712	**0.2349**	0.6390	0.5858
Content-based Model	NIM_S	0.5694	0.2346	0.6272	0.6002
	NIM_W	0.6274	0.2236	**0.7355**	0.6019
Content-based Sequential Model	CSIM_S	0.5490	0.1213	0.6188	0.5731
	CSIM_W	**0.6290**	0.2101	**0.7332**	**0.6058**

Table 6.2: Performances on different products.

(c) PlayStation

Method		Accuracy	F1_Neg	F1_Neu	F1_Pos
Value-based Model	Degroot	0.6115	0.0701	0.7510	0.2023
	Flocking	0.2023	0.1037	0.7332	0.1789
	AsLM	0.5707	0.1985	0.7216	0.3571
	Voter	0.5443	0.2481	0.6543	0.4603
Value-based Sequential Model	TIM	0.6667	0.2205	0.7811	0.4705
Content-based Model	NIM_S	0.6653	0.1301	0.7813	**0.5136**
	NIM_W	0.5925	0.2366	0.7583	0.3400
Content-based Sequential Model	CSIM_S	0.5961	0.0930	0.7361	0.2693
	CSIM_W	**0.6848**	**0.2904**	**0.8193**	0.4908

and forms the future sentiment. Besides, we also notice the following findings from the experiments.

First, we find that the sequential model CSIM_W performs better than two non-sequential models NIM_W and NIM_S on all topics. The better results support our assumption that the communication histories affect the formation of one's opinion. The common understanding of the opinion influence process, which only considers the most recently received information is not adequate. The RNN framework provides a good way to track the opinion changes and capture the correlation between the historical communications and the future sentiment.

Second, we also observe that the performance of CSIM_W is better than that of TIM on all topics. The superior performance of CSIM_W demonstrates that the textual content, though barely studied in social media user behaviors, is actually an important indicator to illustrate the formation of their behaviors. It also motivates us to extend the content-based behavior model to the other scenarios on social media in our future work.

Third, we compare the performances of the two proposed models with different prediction strategies based on the CSIM framework. We find that CSIM_W with the opinion word based prediction strategy significantly outperforms the model CSIM_S with sentiment based prediction strategy on three topics. The improvements verify that the summarized sentiment category labels may harm the performance of the learned model. The model CSIM_W guided by the accurate opinion words could better explain opinion formation.

Finally, we also observe the similar results on the NIM framework. NIM_W also performs better than NIM_S on the topic "Samsung Galaxy" and "Xbox". However, on the topic "PlayStation", NIM_W performs worse than the model NIM_S. Compared with NIM_S, which only predicts the sentiment label, NIM_W needs more information to predict the fine-grained opinion words. Based on the data statistics in Table 3.1, the communication is less active on the topic "PlayStation" compared with the other two topics. The relatively low-active communication on "PlayStation" may not provide enough information for NIM_W to predict opinion words accurately and as a result harms the accuracy of sentiment prediction. The results also indicate that the non-sequential model NIM is sensitive to the topics with different user's involvement on opinion word prediction. Compared with NIM, the sequential model CSIM, which consistently achieves the best results on all topics performs more robust.

Effects of Recurrent Units

In Section 6.4, we propose to employ two popular recurrent units to manipulate the historical information. One is the standard recurrent unit (presented in Equation 6.4), which accumulates the whole past histories. The other is the Gated Recurrent Unit, which controls the integration of the long-term and short-term memories (presented in Equation 6.5-6.8). We examine the performances of the two recurrent units by comparing the models CSIM_SR, CSIM_S, CSIM_W and CSIM_WR, where CSIM_SR and CSIM_WR represent the proposed models with a standard recurrent unit. The performances of these models are presented in Table 6.3.

We first compare the two types of recurrent units on the models with the opinion word based prediction strategy CSIM_W and CSIM_WR. CSIM_W with the GRU unit performs better than CSIM_WR. It implies that during the communication, a user does not keep all historical things in the memory. S/he tends to select the information most relevant to her/his opinions and forgets the irrelevant ones. We also find that GRU has the better prediction ability on the topics "Samsung Galaxy" and "Xbox", for the models with the sentiment based prediction strategy CSIM_S and CSIM_SR. However, on the topic "PlayStation", the standard recurrent unit performs better. It is again because that the communication on the topic "PlayStation" is not as active as that on the other two topics. It is difficult for the model to learn good parameters for selecting the relevant information through GRU. Therefore, taking all the past histories into the consideration achieves better results.

Table 6.3: Performances of models with different recurrent units.

(a) Samsung Galaxy

Method	Accuracy	F1_Neg	F1_Neu	F1_Pos
CSIM_SR	0.3892	0.0641	0.4568	0.3473
CSIM_S	**0.6662**	**0.1981**	**0.7389**	**0.6359**
CSIM_WR	0.6827	0.1973	0.7798	0.6075
CSIM_W	**0.7093**	**0.2534**	**0.7976**	**0.6620**

(b) Xbox

Method	Accuracy	F1_Neg	F1_Neu	F1_Pos
CSIM_SR	0.5320	0.1633	0.5802	0.5685
CSIM_S	**0.5694**	**0.2346**	**0.6272**	**0.6002**
CSIM_WR	0.5955	0.1234	0.7168	0.5381
CSIM_W	**0.6290**	**0.2101**	**0.7332**	**0.6058**

(c) PlayStation

Method	Accuracy	F1_Neg	F1_Neu	F1_Pos
CSIM_SR	**0.6619**	0.0738	**0.7791**	**0.4856**
CSIM_S	0.5961	**0.0930**	0.7361	0.2693
CSIM_WR	0.6548	0.1627	0.7924	**0.4401**
CSIM_W	**0.6849**	**0.2904**	**0.8193**	0.4108

Effects of the Length of History

The proposed model CSIM_W employs a large amount of historical information, which may increase the prediction time and not be feasible for real-time prediction. Additionally, opinions are highly associated with the recent news about the product, and the overlong histories are barely considered by people for current opinion decision. Therefore, we investigate the effective length of history for future sentiments

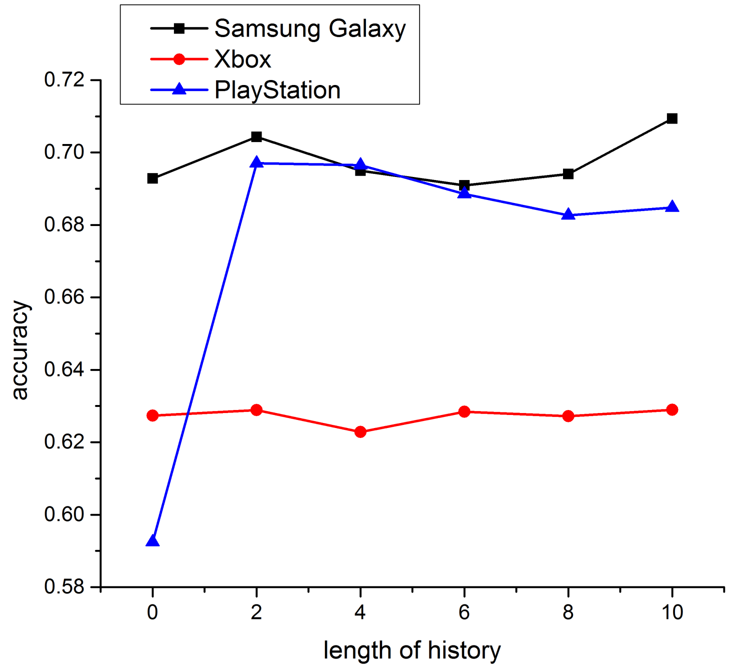


Figure 6.3: Performances on models with different lengths of history

prediction. We compare the model CSIM_W with different lengths of history l, i.e., *0, 2, 4, 6, 8*, where $\boldsymbol{h}_u(t)$ in Equation 6.3 integrates the historical information of previous l timestamps. When the length of history is 0, the model degenerates to NIM_W. The model integrating all historical communication is also compared. The overall accuracy of prediction is plotted in Figure 6.3.

We find that the trends of the accuracy on the lengths of history vary among three topics. When the length of history is 2 timestamps indicating the time interval is 3-7 days, CSIM_W reaches the best performance on the topics "PlayStation" and "Xbox", and second-best performance on the topic "Samsung Galaxy". It reveals that the messages during the past 3-7 days are most important for users to decide their future sentiments. Besides, taking the past 2 timestamps as the history could help significantly reduce the prediction time of the model while at the same time maintaining a high accuracy.

Table 6.4: Performances on opinion word prediction.

(a) Samsung Galaxy

Method	MAP	Pre@5	Pre@10
NIM_W	0.6249	0.6674	0.6996
CSIM_WR	0.6289	0.6662	0.6953
CSIM_W	**0.6467**	**0.6850**	**0.7146**

(b) Xbox

Method	MAP	Pre@5	Pre@10
NIM_W	0.5258	0.5828	0.6262
CSIM_WR	0.4900	0.5526	0.5985
CSIM_W	**0.5314**	**0.5870**	**0.6304**

(c) PlayStation

Method	MAP	Pre@5	Pre@10
NIM_W	0.5795	0.6365	0.6787
CSIM_WR	0.6225	0.6717	0.7082
CSIM_W	**0.6707**	**0.7275**	**0.7651**

6.5.3 Performances of Opinion Word Prediction

In addition to the predictive ability on opinion sentiment, CSIM_W is also capable of predicting the fine-grained opinion information, i.e., opinion words. The opinion words as the elaborated opinion information provide deeper insights into the companies' marketing. To verify the effectiveness of the proposed model on opinion word prediction, we compare CSIM_W with NIM_W, CSIM_WR. All models employ the opinion word based prediction strategy and output a probability distribution over all opinion words in the output layer. We rank the opinion words based on the predicted probabilities in descending order to form a ranking list.

Three evaluation metrics are used to evaluate the performance of opinion word prediction. One is MAP (Mean Average Precision), which has been widely used in evaluating the quality of the ranking list in information retrieval [55].

$$MAP = \frac{\sum_{i=1}^{TN} AveP(i)}{TN}$$

where TN represents the total number of test instances.

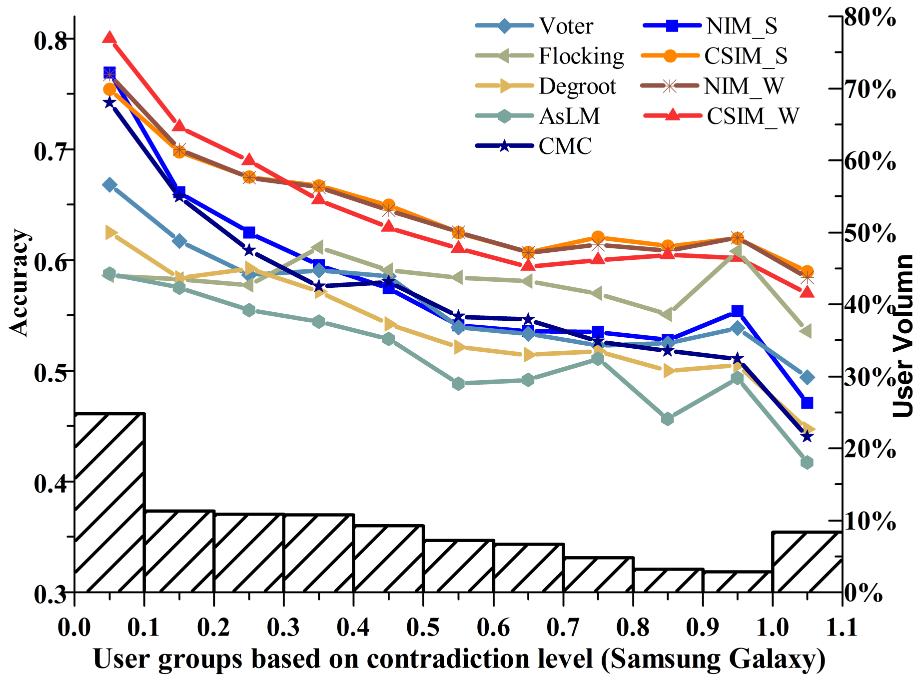
$$AveP(i) = \sum_{j}^{N_i} P(j)/j$$

where $P(j)$ represents the ratio of correctly predicted opinion words up to the position j in the opinion word ranking list, and N_i is the number of opinion words in the gold annotation of the ith test instance. Other two evaluation measurements pre@5 and pre@10 are also considered, which calculate the precision of the first 5 and 10 opinion words in the ranking list respectively. The results are presented in Table 6.4.

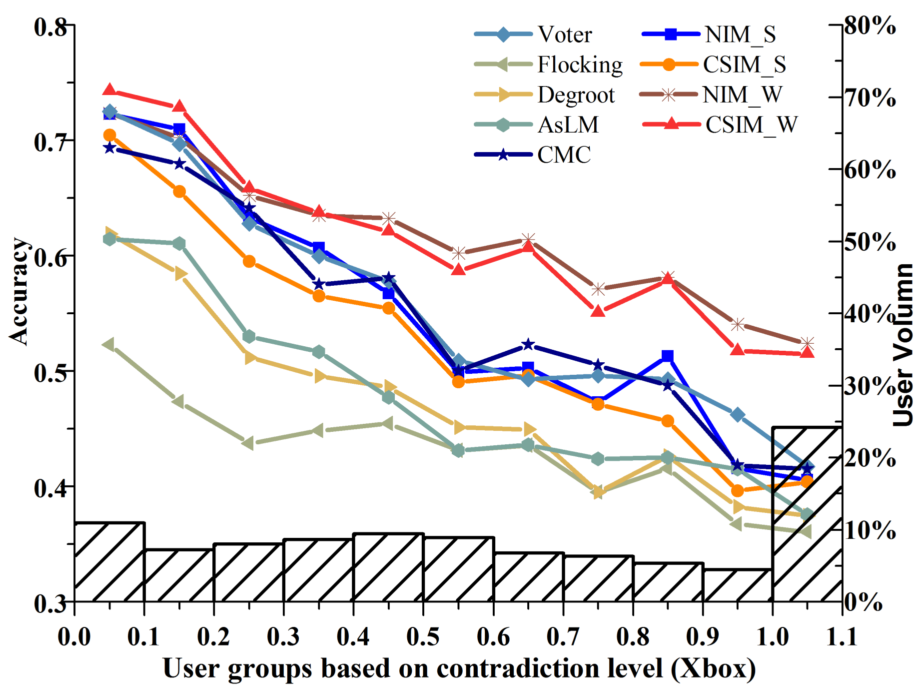
We can see that the CSIM_W model consistently outperforms the other two compared models on all topics in terms of all the evaluation metrics. It demonstrates that the historical information is useful to improve the performance of opinion word prediction, and GRU has a better ability to record the historical information related to one's future opinion expressions.

6.5.4 Performance on Users with Different Sentiment Contradiction Level

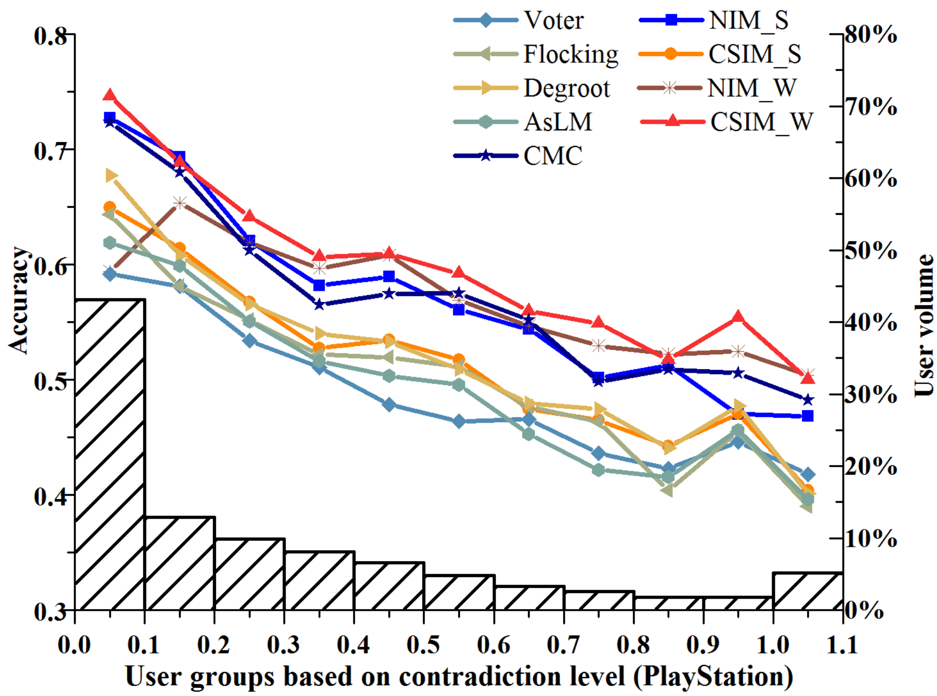
On social media, the characteristics of users' opinion behaviors can be very different. Some users insist on their original opinions and hardly change their opinions. In contrast, some change their previous opinions frequently. Therefore, we further conduct experiments to explore the performances of the models considering users with different characteristics. The sentiment contradiction of all one's tweets is introduced to describe her/his characteristics. It depicts one's belief in the product. For a user who has a firm opinion towards the product, her/his sentiment often remains stable over



(a) Samsung Galaxy



(b) Xbox



(c) PlayStation

Figure 6.3: Performance on users with different sentiment contradiction levels

the time, while for a user who has no fixed opinion, s/he may be easily influenced to change opinion during the communication. We evaluate the performances of our proposed model on smaller groups of users with different sentiment contradiction levels. The sentiment contradiction is measured by considering the mean μ and the variance σ^2 of all the sentiment scores of one's all tweets [161]. It is defined as follows.

$$SC_u = \frac{v{\sigma_u}^2}{v + {\mu_u}^2} W_u$$

where v is a normalization constant, and W_u is a weight function to compensate the contradiction value for users' varying number of tweets.

$$W_u = (1 + exp(\frac{\overline{m(u)} - m(u)}{\beta}))^{-1}$$

where $m(u)$ is the number of u's tweets, $\overline{m(u)}$ is the average number of all users' tweets, and β is a scaling factor. We set v=1 and β=50. Figure 6.3 is a composite of user volume (bars) and prediction accuracy (colored lines) with different sentiment contradictions (x-axis). Users are grouped according to the scores of their sentiment contradictions. For example, the bar between 0 and 0.1 represents the percent of users with the contradiction between 0 and 0.1. For each user group, the prediction accuracy of all the methods on this group is displayed via colored lines.

In Figure 6.3, we observe that both CSIM_W (in red line) and NIM_W (in brown line) perform better than other compared methods on most user groups, especially on the groups of users who change their opinions frequently. We also have an interesting observation. CSIM_W taking past history into account performs better than NIM_W on the users with the low contradiction level, but slightly worse than NIM_W on the users with the high contradiction level. Perhaps the reason is that the easily changed users with the high contradiction level would change their ideas instantly according to the most recently received messages, while the users who already have a fixed opinion need to consider more historical information before changing their original beliefs.

6.5.5 Effects of Data Partitioning

All the above reported results are obtained on the data partition where the ratio of training and test data is 9:1. To demonstrate the robustness of the proposed model, we perform the experiments on another two data partitions where the ratio of training and test data is 8:2 and 7:3 accordingly. The prediction accuracies on the three topics are listed in Figure 6.4. On the topic Xbox and PlayStation, the prediction accuracy does not show a significant difference over different data partitions. On the topic Samsung Galaxy, the prediction accuracy increases when the volume of training data increases. It may indicate that the model could have a better predictability

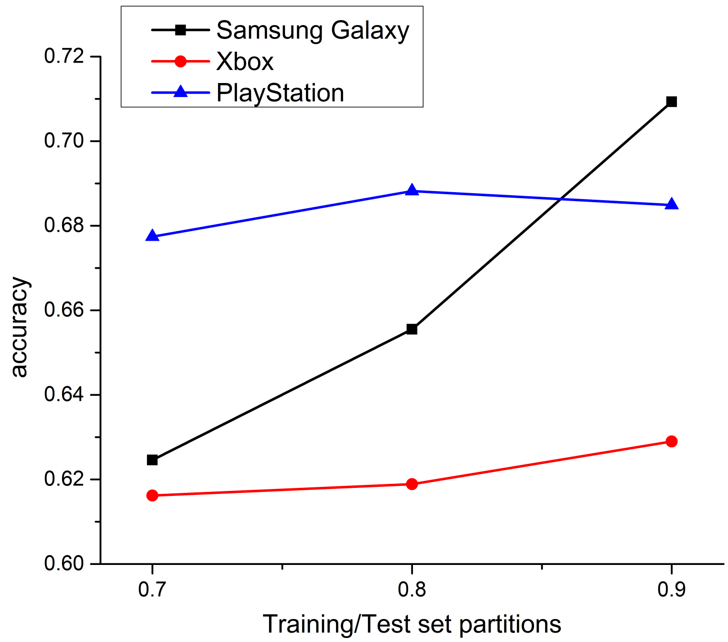


Figure 6.4: Performances on different data partitions

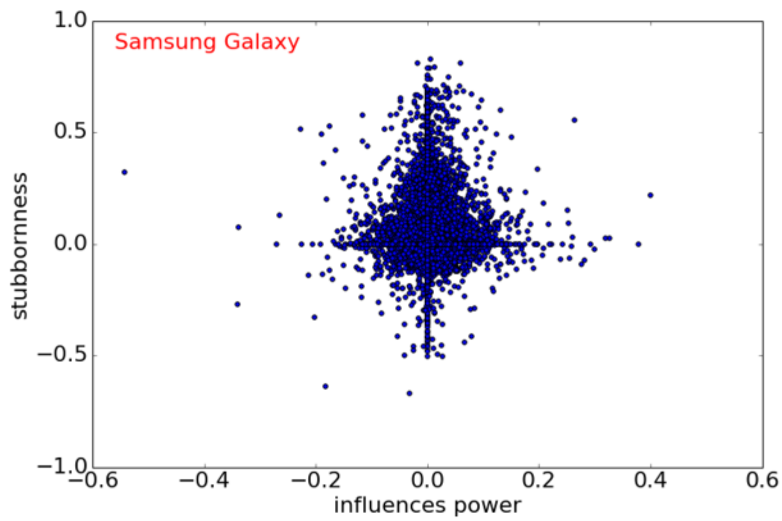
when the volume of training data increases, and when the model has already learned the mechanisms of opinion dynamics, its prediction ability will be stable with time evolving.

6.6 Further Discussion

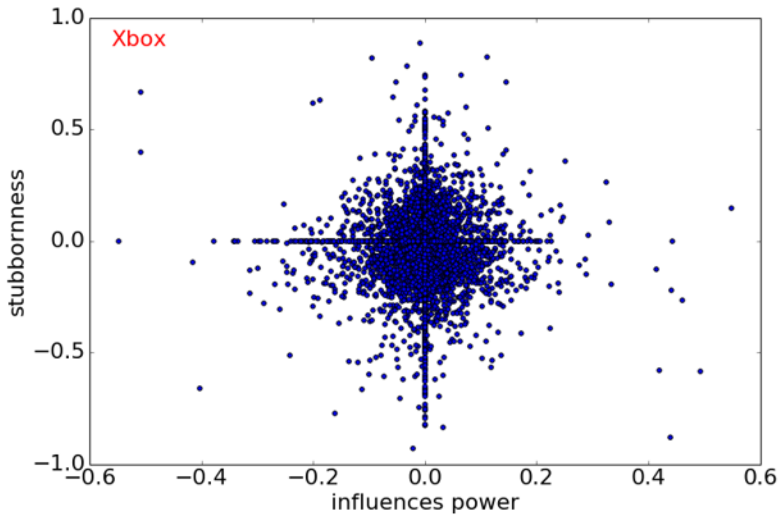
Correlation between Stubbornness and Influence Power

By learning opinion behaviors of users in social network with CSIM, we obtain two factors related to personal characteristics from the learned influence parameters. One is stubbornness α_{u0}, which represents one's insistence on a fixed opinion during the communication. The other is the social influence power of an individual over the whole network, which is measured by averaging the interpersonal influence strengths in the parameter vector $\boldsymbol{\alpha}_{\cdot\boldsymbol{u}}$. To further understand the correlation between one's stubbornness and influence power, we plot each user in Figure 6.4, where the x-axis represents the strength of influence and y-axis represents the strength of stubbornness.

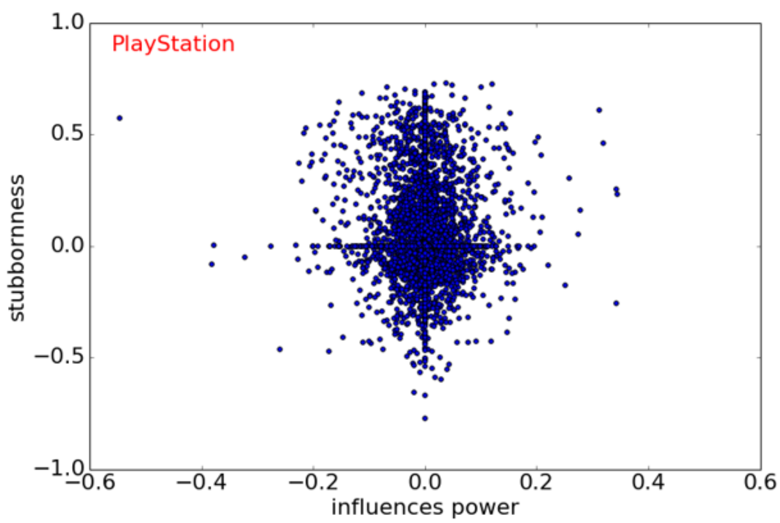
We observe some interesting findings. For users with strong influence power, their stubbornness is relatively low. Meanwhile, users who have high stubbornness have little influence power on her/his neighbors. This can be explained that a user with strong influence evaluates the product from an objective perspective so that her/his opinions will be updated according to the information that s/he obtains. Similarly, users with strong stubbornness usually have fixed opinion and their messages may not provide much information to their followers. Hence their influence is correspondingly weak.



(a) Samsung Galaxy



(b) Xbox



(c) PlayStation

Figure 6.4: Correlation between stubbornness of personality and influence power

6.7 Chapter Summary

In this paper, we introduce a content-based sequential opinion influence framework to incorporate the content information with the historical information for sentiment prediction. Two models with different prediction strategies are proposed. In the experiments, we compare three proposed models and conduct a detailed analysis of their results in terms of the topics with different characteristics. The experimental results demonstrate the effectiveness of the two proposed models. The prediction ability of the proposed model is further verified in the opinion word prediction task. Based on the learned influence, we also find the strong correlation between one's stubbornness and influence power.

Chapter 7

Conclusions and Suggestions for Future Research

The growth of social media over the last decade has revolutionized the way individuals interact with each other and companies conduct business. Understanding and processing the data generated in social media presents challenges and opportunities for a variety of interdisciplinary research, novel algorithms, and tool development. In the area of social media mining, we are endowed with the capability of observing, modeling and evaluating the opinion influence process. Social influence is a valuable treasure to understand human activities on social media. It especially benefits companies for advertisement dissemination, optimization of product impact and other business intelligence related applications. With the techniques and knowledge of computer science, psychology and marketing, we obtain a comprehensive understanding of the underlying mechanism behind the opinion influence process on social media. In the meanwhile, the opinion dynamics, as the result of the opinion influence process, is also tracked and predicted by the proposed models.

In this thesis, we carefully study the problem of opinion influence from data collection, to opinion influence verification and to computational modeling. Previously opinion influence has been studied in the laboratory experiments devised and controlled by researchers, which cannot reflect the real situation. The arise of social media provides researchers opportunities to study the opinion influence process from an empirical view. We collect users' opinion traces from a popular social media platform, i.e., Twitter. Based on the collected data, we demonstrate the existence of opinion influence through a statistical test. Furthermore, a comprehensive study on opinion influence is conducted by exploring three components, including user interactions, temporal dynamics and the exchanged content information. Accordingly, we present three chapters to individually and collectively study these components and demonstrate the effectiveness of the proposed models with empirical experiments. Chapter 4 investigates the temporal dynamics of user interactions, and proposes a temporal opinion influence model to uncover interpersonal opinion influence by capturing interactions between users' opinion dynamics. Chapter 5 focuses on the textual content of the exchanged messages in the opinion influence process, and presents an efficient way to integrate the user identities into the opinion influence modeling. The NN techniques provide a creative way to utilize the semantic information and also make it possible to jointly detect user identity and estimate the opinion influence in a unified model. Finally, Chapter 6 studies both temporal dynamics of opinion behaviors and content of exchanged messages in the opinion influence process upon a sequential NN framework. The different properties of user behaviors and opinion influence in terms of different products are also analyzed through a series of experiments.

We believe that, other than opinion dynamics, many complex dynamics of human society, from buying behaviors in business to voting behaviors in politics, can be better understood by the investigation of the influence process. Specifically, the exploration of the content information, which are mostly expressed in the form of textual messages, provides a more detailed and comprehensive understanding on social media behaviors. The estimation of opinion influence helps us build a better intelligent system and bring advantages to a variety of applications including recommendations, online advertisements, and social media analytics.

7.1 Summary of Contributions

The following sections summarize the contributions of each chapter.

7.1.1 Opinion Influence Test

- We spend 8 months to collect the Twitter data and construct an empirical dataset for studying the opinion influence process. A number of products with different characteristics are studied. For each product, the dataset contains the communication records of involved participant users and their network structure.

- Because the existence of opinion influence on social media has not been verified, we propose a shuffle test approach to verify it. The results of the statistical test prove that the users are indeed influenced by their neighbors' opinions.

7.1.2 Temporal Opinion Influence Modeling

- We first propose to study opinion influence from dynamic user interactions. Accordingly, we develop a temporal opinion influence to track each user's opinion dynamics. Two temporal assumptions are incorporated. Opinion influence is captured by correlating opinion dynamics of different users. Two indicators which capture the effects of social interactions on opinion formation are incorporated.

- The experimental results show that the temporal opinion influence model performs better than the other opinion influence models. It demonstrates that understanding temporal dynamics of user interactions is important and effective for opinion influence modeling.

- Compared with the existing methods that characterize social influence based only on the social network properties, the interpersonal influence learned from the communication records better reflects the actual relationships between users.

- We also observe that opinion influence helps a lot in negative opinion prediction. Compared with positive image management, dealing with negative image is more important for enterprises, which could benefit a lot from opinion influence modeling.

7.1.3 Content-based Opinion Influence Modeling

- Though the User-Generated Text (UGT) is a type of important information on social media, the investigation of the content information is not well considered in opinion influence modeling. We are the first to propose a new content-based opinion influence framework. It explores how the content information affects formation of users' future opinions.

- The proposed model is based on the word embedding techniques and the NN framework. Its capability of capturing the semantic information of textual words and predicting the future sentiment has been verified in the experiments.

- The experiments also demonstrate that the content information greatly improves the accuracy of sentiment prediction compared to the state-based opinion influence model. It indicates that the content-based model better understands the mechanism behind the opinion influence process.

- We observe that utilizing the fact words would bring people positive influence, while emotional words would make people have negative influence towards their followers. These observations are very useful for companies to manage their social media accounts.

- In addition to individually exploring the exchanged content, we also connect users' personalities with their behavior formation. More specifically, we consider user personal identity and social identity in opinion influence modeling.

- A novel framework is proposed to jointly model human behaviors and their social identities in a unified framework. Currently, our social role detection approach is clustering-based. We believe that a supervised or semi-supervised classification can also apply when the necessary data is available.

- The experiments prove that user dual identity can better capture users' characteristics on social media, and the joint learning framework can effectively incorporate user dual identity into opinion influence modeling.

- The combination of dual identities allows the model to better "understand" negative opinion formation when the communication is insufficient.

7.1.4 Content-based Sequential Opinion Influence Modeling

- We propose a novel content-based sequential framework to combine the features explored in the previous studies. The advantages are its abilities to utilize the content information and integrate the history of communication for prediction, which can be further extended to model the content-based user dynamics in other scenarios.

- RNN provides a good way to model the temporal opinion influence process. It has a great ability to not only track the opinion changes but also utilize the semantic information embedded in the message.

- We propose a novel opinion word prediction strategy, whereby the model tries to predict the opinion words first, yielding the sentiment of the opinion. The fine-grained prediction strategy provides a more accurate way for effective model learning and helps companies to make better decisions.

- The performances of the models that separately consider the temporal dynamics and content information vary on the topics with different characteristics. The content-based sequential model best captures the properties of the opinion influence process, and achieves the stable performances on all topics.

- The users who change their opinions frequently are influenced by the most recently received information, while users who already have a fixed opinion need digest more historical information before changing their original beliefs.

- We find that the information in past 3-7 days is most important for decision making.

- An interesting observation about the correlation between a user's stubbornness and influence power is found. A user with a strong influence tends to have a low stubbornness because people would like to trust those who evaluate the product from an objective perspective not from a personal perspective.

- Sentiment prediction is very valuable for business intelligence. The prediction results could help companies to find their target customers in the marketing activities. Considering the huge market potential, even a small improvement on prediction performance can lead to a big improvement in the revenue.

7.2 Future Work

At last, we present the potential extensions of the presented work. Currently, we study opinion influence towards the general features of the product. In fact, a product has a variety of specific features. For example, when people discuss the product

"Samsung Galaxy", they may concern different aspects of it, such as "battery", "screen", and "design", etc. Understanding the fine-grained opinions included in the messages could help us better capture opinion influence, and also provide opportunities for a more precise prediction, i.e., predicting users' opinions towards certain particular aspects. In the future study, we plan to conduct a fine-grained opinion influence study.

Several problems need to be solved in the fine-grained opinion influence analysis. First, how to understand aspect-based opinions conveyed in the messages? Extracting the pair of aspect and opinion has been extensively studied by researchers [157, 98, 96]. Recently, some studies propose to solve the problem of aspect sentiment classification by employing NN techniques [166, 154]. These techniques make it possible to combine the aspect-based opinion into our NN-based opinion influence modeling framework. Second, how to distinguish opinion influence in different aspects? The most straightforward way is to assume the influence in different aspects to be equal. However, it does not always reflect the fact. A more reasonable way is to capture the different influence weights realted to different aspects and explore their effects in the opinion influence process. The last challenge is how to predict aspect-based opinion. The model should have the ability to predict a pair of aspect and sentiment at the same time. Besides, a person can express her/his opinion on more than one product feature in the same message. For example, "I like the design of Samsung Galaxy S8, but the battery life is short". The multiple prediction is also needed. An extensive study has been conducted on multi-class classification, and the details of the studies can be found in the survey [160]. These techniques provide a lot of references for us when developing the multiple aspect-opinion prediction. All the above-proposed problems are expected to study in the future work.

Bibliography

[1] Daron Acemoglu, Asuman Ozdaglar, and Ali ParandehGheibi. Spread of (mis) information in social networks. *Games and Economic Behavior*, 70(2):194–227, 2010.

[2] Charu C Aggarwal and ChengXiang Zhai. A survey of text classification algorithms. *Mining text data*, pages 163–222, 2012.

[3] Alan Agresti. *Logistic regression*. Wiley Online Library, 2002.

[4] Aris Anagnostopoulos, Ravi Kumar, and Mohammad Mahdian. Influence and correlation in social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 7–15. ACM, 2008.

[5] Theodore W Anderson and Leo A Goodman. Statistical inference about markov chains. *The Annals of Mathematical Statistics*, pages 89–110, 1957.

[6] Chalee Asavathiratham, Sandip Roy, Bernard Lesieutre, and George Verghese. The influence model. *IEEE Control Systems*, 21(6):52–64, 2001.

[7] Solomon E Asch. Opinions and social pressure. *Scientific American*, 193(5):31–35, 1955.

[8] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In *LREC*, volume 10, pages 2200–2204, 2010.

[9] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[10] Frédéric Bastien, Pascal Lamblin, Razvan Pascanu, James Bergstra, Ian Goodfellow, Arnaud Bergeron, Nicolas Bouchard, David Warde-Farley, and Yoshua Bengio. Theano: new features and speed improvements. *arXiv preprint arXiv:1211.5590*, 2012.

[11] Sumit Basu, Tanzeem Choudhury, Brian Clarkson, Alex Pentland, et al. Learning human interactions with the influence model. NIPS, 2001.

[12] Samy Bengio and Yoshua Bengio. Taking on the curse of dimensionality in joint distributions using neural networks. *IEEE Transactions on Neural Networks*, 11(3):550–557, 2000.

[13] Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Janvin. A neural probabilistic language model. *The Journal of Machine Learning Research*, 3:1137–1155, 2003.

[14] Yoshua Bengio, Jean-Sébastien Senécal, et al. Quick training of probabilistic neural nets by importance sampling. In *AISTATS*, 2003.

[15] Yoshua Bengio, Patrice Simard, and Paolo Frasconi. Learning long-term dependencies with gradient descent is difficult. *Neural Networks, IEEE Transactions on*, 5(2):157–166, 1994.

[16] Adam L Berger, Vincent J Della Pietra, and Stephen A Della Pietra. A maximum entropy approach to natural language processing. *Computational linguistics*, 22(1):39–71, 1996.

[17] James C Bezdek. Objective function clustering. In *Pattern recognition with fuzzy objective function algorithms*, pages 43–93. Springer, 1981.

[18] Adrian Bickerstaffe and Ingrid Zukerman. A hierarchical classifier applied to multi-way sentiment detection. In *Proceedings of the 23rd international conference on computational linguistics*, pages 62–70. Association for Computational Linguistics, 2010.

[19] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[20] Stephen P Borgatti and Martin G Everett. Notions of position in social network analysis. *Sociological methodology*, pages 1–35, 1992.

[21] Christian Borgs, Michael Brautbar, Jennifer Chayes, and Brendan Lucier. Maximizing social influence in nearly optimal time. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 946–957. SIAM, 2014.

[22] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[23] Matthew Brand, Nuria Oliver, and Alex Pentland. Coupled hidden markov models for complex action recognition. In *Computer vision and pattern recognition, 1997. proceedings., 1997 ieee computer society conference on*, pages 994–999. IEEE, 1997.

[24] Phil E Brown and Junlan Feng. Measuring user influence on twitter using modified k-shell decomposition. In *Fifth international AAAI conference on weblogs and social media*, 2011.

[25] Ronald S Burt. Detecting role equivalence. *Social networks*, 12(1):83–97, 1990.

[26] Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, and P Krishna Gummadi. Measuring user influence in twitter: The million follower fallacy. *Icwsm*, 10(10-17):30, 2010.

[27] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[28] Bernard Chazelle. Natural algorithms and influence systems. *Communications of the ACM*, 55(12):101–110, 2012.

[29] Chengyao Chen, Dehong Gao, Wenjie Li, and Yuexian Hou. Inferring topic-dependent influence roles of twitter users. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 1203–1206. ACM, 2014.

[30] Chengyao Chen, Wenjie Li, Dehong Gao, and Yuexian Hou. Exploring interpersonal influence by tracking user dynamic interactions. *IEEE Intelligent Systems*, 32(3):28–35, 2017.

[31] Chengyao Chen, Zhitao Wang, Yu Lei, and Wenjie Li. Content-based influence modeling for opinion behavior prediction. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 2207–2216, 2016.

[32] Chengyao Chen, Zhitao Wang, and Wenjie Li. Modeling opinion influence with user dual identity. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 2019–2022. ACM, 2017.

[33] Chengyao Chen, Zhitao Wang, and Wenjie Li. Tracking dynamics of opinion behaviors with a content-based sequential opinion influence model. *IEEE Transactions on Affective Computing*, 2018.

[34] Wei Chen, Yajun Wang, and Siyu Yang. Efficient influence maximization in social networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 199–208. ACM, 2009.

[35] Wei Chen, Yifei Yuan, and Li Zhang. Scalable influence maximization in social networks under the linear threshold model. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, pages 88–97. IEEE, 2010.

[36] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

[37] Edward Choi, Mohammad Taha Bahadori, Elizabeth Searles, Catherine Coffey, Michael Thompson, James Bost, Javier Tejedor-Sojo, and Jimeng Sun. Multi-layer representation learning for medical concepts. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1495–1504. ACM, 2016.

[38] George Christodoulides, Colin Jevons, and Jennifer Bonhomme. Memo to marketers: Quantitative evidence for change: How user-generated content really affects brands. *Journal of advertising research*, 52(1):53–64, 2012.

[39] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.

[40] Peter Clifford and Aidan Sudbury. A model for spatial conflict. *Biometrika*, 60(3):581–588, 1973.

[41] Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[42] Peng Cui, Fei Wang, Shaowei Liu, Mingdong Ou, Shiqiang Yang, and Lifeng Sun. Who should share what?: item-level social influence prediction for users and posts ranking. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 185–194. ACM, 2011.

[43] Abhimanyu Das, Sreenivas Gollapudi, and Kamesh Munagala. Modeling opinion dynamics in social networks. In *Proceedings of the 7th ACM international conference on Web search and data mining*, pages 403–412. ACM, 2014.

[44] Abir De, Sourangshu Bhattacharya, Parantapa Bhattacharya, Niloy Ganguly, and Soumen Chakrabarti. Learning a linear influence model from transient opinion dynamics. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 401–410. ACM, 2014.

[45] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. Predicting depression via social media. *ICWSM*, 13:1–10, 2013.

[46] Morris H DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.

[47] Morton Deutsch and Harold B Gerard. A study of normative and informational social influences upon individual judgment. *The journal of abnormal and social psychology*, 51(3):629, 1955.

[48] Xiaowen Ding, Bing Liu, and Philip S Yu. A holistic lexicon-based approach to opinion mining. In *Proceedings of the 2008 international conference on web search and data mining*, pages 231–240. ACM, 2008.

[49] Pedro Domingos and Matt Richardson. Mining the network value of customers. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 57–66. ACM, 2001.

[50] Joseph C Dunn. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. 1973.

[51] Jubin Edachery, Arunabha Sen, and Franz J Brandenburg. Graph clustering using distance-k cliques. In *International Symposium on Graph Drawing*, pages 98–106. Springer, 1999.

[52] Jeffrey L Elman. Finding structure in time. *Cognitive science*, 14(2):179–211, 1990.

[53] Katherine Faust. Comparison of methods for positional analysis: Structural and general equivalences. *Social Networks*, 10(4):313–341, 1988.

[54] Kristine L Fitch and Robert E Sanders. *Handbook of language and social interaction*. Psychology Press, 2004.

[55] William Bruce Frakes and Ricardo Baeza-Yates. *Information retrieval: Data structures & algorithms*, volume 331. prentice Hall Englewood Cliffs, New Jersey, 1992.

[56] Noah E Friedkin and Eugene C Johnsen. Social influence and opinions. *Journal of Mathematical Sociology*, 15(3-4):193–206, 1990.

[57] Nir Friedman, Dan Geiger, and Moises Goldszmidt. Bayesian network classifiers. *Machine learning*, 29(2-3):131–163, 1997.

[58] Serge Galam. Minority opinion spreading in random geometry. *The European Physical Journal B-Condensed Matter and Complex Systems*, 25(4):403–406, 2002.

[59] Cynthia Gallois. *The handbook of intergroup communication*. Routledge, 2012.

[60] Javad Ghaderi and R Srikant. Opinion dynamics in social networks with stubborn agents: Equilibrium and convergence rate. *Automatica*, 50(12):3209–3215, 2014.

[61] Jacob Goldenberg, Barak Libai, and Eitan Muller. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing letters*, 12(3):211–223, 2001.

[62] Manuel Gomez Rodriguez, Jure Leskovec, and Andreas Krause. Inferring networks of diffusion and influence. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1019–1028. ACM, 2010.

[63] Ward Hunt Goodenough. *Property, kin, and community on Truk*. Shoe String Pr Inc, 1978.

[64] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.

[65] Amit Goyal, Francesco Bonchi, and Laks VS Lakshmanan. A data-based approach to social influence maximization. *Proceedings of the VLDB Endowment*, 5(1):73–84, 2011.

[66] Amit Goyal, Wei Lu, and Laks VS Lakshmanan. Celf++: optimizing the greedy algorithm for influence maximization in social networks. In *Proceedings of the 20th international conference companion on World wide web*, pages 47–48. ACM, 2011.

[67] Amit Goyal, Wei Lu, and Laks VS Lakshmanan. Simpath: An efficient algorithm for influence maximization under the linear threshold model. In *Data Mining (ICDM), 2011 IEEE 11th International Conference on*, pages 211–220. IEEE, 2011.

[68] Mark Granovetter. Threshold models of collective behavior. *American journal of sociology*, 83(6):1420–1443, 1978.

[69] Lin Gui, Ruifeng Xu, Yulan He, Qin Lu, and Zhongyu Wei. Intersubjectivity and sentiment: From language to knowledge. In *IJCAI*, pages 2789–2795, 2016.

[70] Lin Gui, Yu Zhou, Ruifeng Xu, Yulan He, and Qin Lu. Learning representations from heterogeneous network for sentiment classification of product reviews. *Knowledge-Based Systems*, 124:34–45, 2017.

[71] Adrien Guille, Hakim Hacid, Cecile Favre, and Djamel A Zighed. Information diffusion in online social networks: A survey. *ACM Sigmod Record*, 42(2):17–28, 2013.

[72] Meng Han, Mingyuan Yan, Zhipeng Cai, Yingshu Li, Xingquan Cai, and Jiguo Yu. Influence maximization by probing partial communities in dynamic online social networks. *Transactions on Emerging Telecommunications Technologies*, 28(4), 2017.

[73] Jeffrey T Hancock, Christopher Landrigan, and Courtney Silver. Expressing emotion in text-based communication. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 929–932. ACM, 2007.

[74] Frank Harary. A criterion for unanimity in french's theory of social power. 1959.

[75] Rainer Hegselmann and Ulrich Krause. Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation*, 5(3), 2002.

[76] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[77] Richard A Holley and Thomas M Liggett. Ergodic theorems for weakly interacting infinite systems and the voter model. *The annals of probability*, pages 643–663, 1975.

[78] Baotian Hu, Zhengdong Lu, Hang Li, and Qingcai Chen. Convolutional neural network architectures for matching natural language sentences. In *Advances in neural information processing systems*, pages 2042–2050, 2014.

[79] Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177. ACM, 2004.

[80] Xia Hu and Huan Liu. Social status and role analysis of palin's email network. In *Proceedings of the 21st International Conference on World Wide Web*, pages 531–532. ACM, 2012.

[81] Peter Hui and Michelle Gregory. Quantifying sentiment and influence in blogspaces. In *Proceedings of the First Workshop on Social Media Analytics*, pages 53–61. ACM, 2010.

[82] CJ Hutto and Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth International AAAI Conference on Weblogs and Social Media*, 2014.

[83] Clayton J Hutto, Sarita Yardi, and Eric Gilbert. A longitudinal study of follow predictors on twitter. In *Proceedings of the sigchi conference on human factors in computing systems*, pages 821–830. ACM, 2013.

[84] Nina Farisha Isa, Noor Akma Mohd Salleh, and Azmin Azliza Aziz. Determinants and impact of online social interaction on online buying behaviour. *Procedia-Social and Behavioral Sciences*, 219:352–358, 2016.

[85] Xin Jin and Yaohua Wang. Research on social network structure and public opinions dissemination of micro-blog based on complex network analysis. *JNW*, 8(7):1543–1550, 2013.

[86] Andreas M Kaplan and Michael Haenlein. Users of the world, unite! the challenges and opportunities of social media. *Business horizons*, 53(1):59–68, 2010.

[87] David Kempe, Jon Kleinberg, Sigal Oren, and Aleksandrs Slivkins. Selection and influence in cultural dynamics. *Network Science*, 4(1):1–27, 2016.

[88] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM, 2003.

[89] Alistair Kennedy and Diana Inkpen. Sentiment classification of movie reviews using contextual valence shifters. *Computational intelligence*, 22(2):110–125, 2006.

[90] Jinha Kim, Seung-Keol Kim, and Hwanjo Yu. Scalable and parallelizable processing of influence maximization for large-scale social networks? In *Data Engineering (ICDE), 2013 IEEE 29th International Conference on*, pages 266–277. IEEE, 2013.

[91] Soo-Min Kim and Eduard Hovy. Determining the sentiment of opinions. In *Proceedings of the 20th international conference on Computational Linguistics*, page 1367. Association for Computational Linguistics, 2004.

[92] Yoon Kim. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*, 2014.

[93] Masahiro Kimura and Kazumi Saito. Tractable models for information diffusion in social networks. *Knowledge Discovery in Databases: PKDD 2006*, pages 259–271, 2006.

[94] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[95] Jon M Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632, 1999.

[96] Nozomi Kobayashi, Kentaro Inui, and Yuji Matsumoto. Extracting aspect-evaluation and aspect-of relations in opinion mining. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, 2007.

[97] PL Krapivsky and S Redner. Dynamics of majority rule in two-state interacting spin systems. *Physical Review Letters*, 90(23):238701, 2003.

[98] Lun-Wei Ku, Yu-Ting Liang, and Hsin-Hsi Chen. Opinion extraction, summarization and tracking in news and blog corpora. In *Proceedings of AAAI*, pages 100–107, 2006.

[99] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, pages 591–600. ACM, 2010.

[100] Quoc V Le and Tomas Mikolov. Distributed representations of sentences and documents. *arXiv preprint arXiv:1405.4053*, 2014.

[101] Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie Glance. Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 420–429. ACM, 2007.

[102] Anton Leuski. Email is a stage: discovering people roles from email archives. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 502–503. ACM, 2004.

[103] Omer Levy and Yoav Goldberg. Dependency-based word embeddings. In *ACL (2)*, pages 302–308, 2014.

[104] Jiwei Li and Dan Jurafsky. Do multi-sense embeddings improve natural language understanding? *arXiv preprint arXiv:1506.01070*, 2015.

[105] Jiwei Li, Will Monroe, Tianlin Shi, Alan Ritter, and Dan Jurafsky. Adversarial learning for neural dialogue generation. *arXiv preprint arXiv:1701.06547*, 2017.

[106] Shoushan Li, Sophia Yat Mei Lee, Ying Chen, Chu-Ren Huang, and Guodong Zhou. Sentiment classification and polarity shifting. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 635–643. Association for Computational Linguistics, 2010.

[107] Renjie Liao, Alex Schwing, Richard Zemel, and Raquel Urtasun. Learning deep parsimonious representations. In *Advances in Neural Information Processing Systems*, pages 5076–5084, 2016.

[108] Pengfei Liu, Shafiq Joty, and Helen Meng. Fine-grained opinion mining with recurrent neural networks and word embeddings. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1433–1443, 2015.

[109] Charles G Lord, Lee Ross, and Mark R Lepper. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of personality and social psychology*, 37(11):2098, 1979.

[110] Minh-Thang Luong, Hieu Pham, and Christopher D Manning. Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025*, 2015.

[111] Sonia Martínez, Francesco Bullo, Jorge Cortés, and Emilio Frazzoli. On synchronous robotic networkspart i: Models, tasks, and complexity. *IEEE Transactions on Automatic Control*, 52(12):2199–2213, 2007.

[112] Andrew McCallum, Xuerui Wang, and Andrés Corrada-Emmanuel. Topic and role discovery in social networks with experiments on enron and academic email. *Journal of Artificial Intelligence Research*, 30:249–272, 2007.

[113] Ryan McDonald, Kerry Hannan, Tyler Neylon, Mike Wells, and Jeff Reynar. Structured models for fine-to-coarse sentiment analysis. In *Annual meeting-association for computational linguistics*, volume 45, page 432, 2007.

[114] Walaa Medhat, Ahmed Hassan, and Hoda Korashy. Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4):1093–1113, 2014.

[115] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[116] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.

[117] Alan Mislove, Bimal Viswanath, Krishna P Gummadi, and Peter Druschel. You are who you know: inferring user profiles in online social networks. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 251–260. ACM, 2010.

[118] Richard W Mizerski. An attribution explanation of the disproportionate influence of unfavorable information. *Journal of Consumer Research*, 9(3):301–310, 1982.

[119] Frederic Morin and Yoshua Bengio. Hierarchical probabilistic neural network language model. In *Aistats*, volume 5, pages 246–252, 2005.

[120] Mehdi Moussaid, Simon Garnier, Guy Theraulaz, and Dirk Helbing. Collective information processing and pattern formation in swarms, flocks, and crowds. *Topics in Cognitive Science*, 1(3):469–497, 2009.

[121] Mehdi Moussaïd, Juliane E Kämmer, Pantelis P Analytis, and Hansjörg Neth. Social influence and the collective dynamics of opinion formation. *PloS one*, 8(11):e78433, 2013.

[122] Tony Mullen and Nigel Collier. Sentiment analysis using support vector machines with diverse information sources. In *EMNLP*, volume 4, pages 412–418, 2004.

[123] Seth Myers and Jure Leskovec. On the convexity of latent social network inference. In *Advances in neural information processing systems*, pages 1741–1749, 2010.

[124] Jin-Cheon Na, Haiyang Sui, Christopher SG Khoo, Syin Chan, and Yunyun Zhou. Effectiveness of simple linguistic processing in automatic sentiment classification of product reviews. 2004.

[125] Rinkesh Nagmoti, Ankur Teredesai, and Martine De Cock. Ranking approaches for microblog search. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*, volume 1, pages 153–157. IEEE, 2010.

[126] Pradeep Natarajan and Ramakant Nevatia. Coupled hidden semi markov models for activity recognition. In *Motion and Video Computing, 2007. WMVC'07. IEEE Workshop on*, pages 10–10. IEEE, 2007.

[127] Yurii Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2013.

[128] Tomoya Noro, Fei Ru, Feng Xiao, and Takehiro Tokuda. Twitter user rank using keyword search. *Information Modelling and Knowledge Bases XXIV. Frontiers in Artificial Intelligence and Applications*, 251:31–48, 2013.

[129] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.

[130] Aditya Pal and Scott Counts. Identifying topical authorities in microblogs. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 45–54. ACM, 2011.

[131] Pingbo Pan, Zhongwen Xu, Yi Yang, Fei Wu, and Yueting Zhuang. Hierarchical recurrent neural encoder for video representation with application to captioning. *arXiv preprint arXiv:1511.03476*, 2015.

[132] Wei Pan, Wen Dong, Manuel Cebrian, Taemie Kim, James H Fowler, and Alex Sandy Pentland. Modeling dynamical influence in human interaction: Using data to make better inferences about influence within social systems. *IEEE Signal Processing Magazine*, 29(2):77–86, 2012.

[133] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pages 79–86. Association for Computational Linguistics, 2002.

[134] Razvan Pascanu, Caglar Gulcehre, Kyunghyun Cho, and Yoshua Bengio. How to construct deep recurrent neural networks. *arXiv preprint arXiv:1312.6026*, 2013.

[135] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. On the difficulty of training recurrent neural networks. In *International Conference on Machine Learning*, pages 1310–1318, 2013.

[136] James W Pennebaker, Martha E Francis, and Roger J Booth. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71:2001, 2001.

[137] Ana-Maria Popescu and Orena Etzioni. Extracting product features and opinions from reviews. In *Natural language processing and text mining*, pages 9–28. Springer, 2007.

[138] Ariadna Quattoni, Sybor Wang, Louis-Philippe Morency, Morency Collins, and Trevor Darrell. Hidden conditional random fields. *IEEE transactions on pattern analysis and machine intelligence*, 29(10), 2007.

[139] Adithya Rao, Nemanja Spasojevic, Zhisheng Li, and Trevor DSouza. Klout score: Measuring influence across multiple social networks. In *Big Data (Big Data), 2015 IEEE International Conference on*, pages 2282–2289. IEEE, 2015.

[140] Haibing Ren and Guangyou Xu. Human action recognition with primitive-based coupled-hmm. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, volume 2, pages 494–498. IEEE, 2002.

[141] Yafeng Ren, Yue Zhang, Meishan Zhang, and Donghong Ji. Context-sensitive twitter sentiment classification using neural network. In *AAAI*, pages 215–221, 2016.

[142] Manoel Horta Ribeiro, Pedro H Calais, Yuri A Santos, Virgílio AF Almeida, and Wagner Meira Jr. " like sheep among wolves": Characterizing hateful users on twitter. *arXiv preprint arXiv:1801.00317*, 2017.

[143] Manuel Gomez Rodriguez, David Balduzzi, and Bernhard Schölkopf. Uncovering the temporal dynamics of diffusion networks. *arXiv preprint arXiv:1105.0697*, 2011.

[144] Manuel Gomez Rodriguez and Bernhard Schölkopf. Submodular inference of diffusion networks from multiple trees. *arXiv preprint arXiv:1205.1671*, 2012.

[145] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *Cognitive modeling*, 5:3, 1988.

[146] Kazumi Saito, Kouzou Ohara, Yuki Yamagishi, Masahiro Kimura, and Hiroshi Motoda. Learning diffusion probability based on node attributes in social networks. In *International Symposium on Methodologies for Intelligent Systems*, pages 153–162. Springer, 2011.

[147] Iulian Vlad Serban, Alessandro Sordoni, Yoshua Bengio, Aaron C Courville, and Joelle Pineau. Building end-to-end dialogue systems using generative hierarchical neural network models. In *AAAI*, volume 16, pages 3776–3784, 2016.

[148] Javier Serrano-Puche. Herramientas web para la medición de la influencia digital: análisis de klout y peerindex. 2012.

[149] Richard Socher, Cliff C Lin, Chris Manning, and Andrew Y Ng. Parsing natural scenes and natural language with recursive neural networks. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 129–136, 2011.

[150] Jimeng Sun and Jie Tang. A survey of models and algorithms for social influence analysis. *Social network data analytics*, pages 177–214, 2011.

[151] Johan AK Suykens and Joos Vandewalle. Least squares support vector machine classifiers. *Neural processing letters*, 9(3):293–300, 1999.

[152] Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2):267–307, 2011.

[153] Duyu Tang, Bing Qin, and Ting Liu. Document modeling with gated recurrent neural network for sentiment classification. In *EMNLP*, pages 1422–1432, 2015.

[154] Duyu Tang, Bing Qin, and Ting Liu. Aspect level sentiment classification with deep memory network. *arXiv preprint arXiv:1605.08900*, 2016.

[155] Duyu Tang, Furu Wei, Nan Yang, Ming Zhou, Ting Liu, and Bing Qin. Learning sentiment-specific word embedding for twitter sentiment classification. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, volume 1, pages 1555–1565, 2014.

[156] Youze Tang, Yanchen Shi, and Xiaokui Xiao. Influence maximization in near-linear time: A martingale approach. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, pages 1539–1554. ACM, 2015.

[157] Ivan Titov and Ryan McDonald. A joint model of text and aspect ratings for sentiment summarization. *proceedings of ACL-08: HLT*, pages 308–316, 2008.

[158] Karen Tracy and Kathleen Haspel. Language and social interaction: Its institutional identity, intellectual landscape, and discipline-shifting agenda. *Journal of communication*, 54(4):788–816, 2004.

[159] John Nikolas Tsitsiklis. Problems in decentralized decision making and computation. Technical report, MASSACHUSETTS INST OF TECH CAMBRIDGE LAB FOR INFORMATION AND DECISION SYSTEMS, 1984.

[160] Grigorios Tsoumakas and Ioannis Katakis. Multi-label classification: An overview. *International Journal of Data Warehousing and Mining*, 3(3), 2006.

[161] Mikalai Tsytsarau, Themis Palpanas, and Kerstin Denecke. Scalable detection of sentiment-based contradictions. *DiversiWeb, WWW*, 2011, 2011.

[162] Peter D Turney. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 417–424. Association for Computational Linguistics, 2002.

[163] Tracy L Tuten and Michael R Solomon. *Social media marketing*. Sage, 2017.

[164] Jia Wang, Vincent W Zheng, Zemin Liu, and Kevin Chen-Chuan Chang. Topological recurrent neural network for diffusion prediction. *arXiv preprint arXiv:1711.10162*, 2017.

[165] Yaxuan Wang, Hongzhi Wang, Jianzhong Li, and Hong Gao. Efficient influence maximization in weighted independent cascade model. In *International Conference on Database Systems for Advanced Applications*, pages 49–64. Springer, 2016.

[166] Yequan Wang, Minlie Huang, Li Zhao, et al. Attention-based lstm for aspect-level sentiment classification. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 606–615, 2016.

[167] Yongqing Wang, Huawei Shen, Shenghua Liu, Jinhua Gao, and Xueqi Cheng. Cascade dynamics modeling with attention-based recurrent neural network. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pages 2985–2991. AAAI Press, 2017.

[168] Yu Wang, Gao Cong, Guojie Song, and Kunqing Xie. Community-based greedy algorithm for mining top-k influential nodes in mobile social networks. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1039–1048. ACM, 2010.

[169] Howard T Welser, Dan Cosley, Gueorgi Kossinets, Austin Lin, Fedor Dokshin, Geri Gay, and Marc Smith. Finding social roles in wikipedia. In *Proceedings of the 2011 iConference*, pages 122–129. ACM, 2011.

[170] Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He. Twitterrank: finding topic-sensitive influential twitterers. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 261–270. ACM, 2010.

[171] Casey Whitelaw, Navendu Garg, and Shlomo Argamon. Using appraisal groups for sentiment analysis. In *Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 625–631. ACM, 2005.

[172] Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, et al. Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*, 2016.

[173] Caiming Xiong, Stephen Merity, and Richard Socher. Dynamic memory networks for visual and textual question answering. In *International Conference on Machine Learning*, pages 2397–2406, 2016.

[174] Jaewon Yang and Jure Leskovec. Modeling information diffusion in implicit networks. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, pages 599–608. IEEE, 2010.

[175] Yang Yang, Jie Tang, Cane Wing-ki Leung, Yizhou Sun, Qicong Chen, Juanzi Li, and Qiang Yang. Rain: Social role-aware information diffusion. In *AAAI*, pages 367–373, 2015.

[176] Zichao Yang, Xiaodong He, Jianfeng Gao, Li Deng, and Alex Smola. Stacked attention networks for image question answering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 21–29, 2016.

[177] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alexander J Smola, and Eduard H Hovy. Hierarchical attention networks for document classification. In *HLT-NAACL*, pages 1480–1489, 2016.

[178] Ilan Yaniv. Receiving other peoples advice: Influence and benefit. *Organizational Behavior and Human Decision Processes*, 93(1):1–13, 2004.

[179] Ercan Yildiz, Daron Acemoglu, Asuman E Ozdaglar, Amin Saberi, and Anna Scaglione. Discrete opinion dynamics with stubborn agents. *Available at SSRN 1744113*, 2011.

[180] Fabio Massimo Zanzotto, Ioannis Korkontzelos, Francesca Fallucchi, and Suresh Manandhar. Estimating linear models for compositional distributional semantics. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 1263–1271. Association for Computational Linguistics, 2010.

[181] Yuchen Zhao, Neel Sundaresan, Zeqian Shen, and Philip S Yu. Anatomy of a web-scale resale market: a data mining approach. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1533–1544. ACM, 2013.

[182] Yuchen Zhao, Guan Wang, Philip S Yu, Shaobo Liu, and Simon Zhang. Inferring social roles and statuses in social networks. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 695–703. ACM, 2013.