Information diffusion analysis in online social networks based on deep representation learning

Chen, X.

Chapter 2

Literature Review

In this chapter, we provide a brief review of the studies that are relevant to this thesis. We focus on summarizing the existing methods for two specific tasks: information cascade modeling and rumor detection. Existing methods for both tasks can be divided into two groups, i.e., conventional and deep learning-based methods. Since this thesis is inspired by deep learning, our literature review concentrate more on deep learning-based methods and only briefly describe conventional methods.

2.1 Information Cascades Modeling

As we mentioned in Section 1.2, in this thesis, we focus on modeling the information cascade through macro-level information cascade prediction tasks. The macro-level information cascade prediction tasks aim at modeling the cascade scale via estimating the future popularity of the diffusion cascade. The information cascade is a phenomenon caused by information transmission from one user to another based on social interactions (e.g., follower/following) in OSNs, which always used to describe the information diffusion and consists of the trajectories and structures of information diffusion, as well as the participants in information spreading [11].

2.1.1 Conventional methods

Conventional methods in macro-level information cascade prediction mainly fall into two categories: (1) point process-based methods, and (2) handcrafted feature-based methods.

Point process-based methods: Most individual activities in the social system can be described as a point process [47]. The point process-based methods regard message diffusion as the arrival process of users' retweet behavior. Specifically, these methods focus on modeling the intensity function in the arrival process for each message independently, it observes each event within the observation window and learns

the parameters through maximizing the probability of events occurring during a period of time. Typical point process methods include Poisson process [2, 30, 48] and Hawkes process [49, 50]. Both of these point processes are committed to describing the key factors in message diffusion: (1) self-influence–i.e., each retweet user will influence the trend of future retweets, (2) time-decay effect–i.e., the influence of a retweet user decays with the time elapsed, and (3) rich-get-richer phenomenon–i.e., a message shared by influential users will get more retweets. Shen et al. [2] upgraded the Poisson process to reinforced Poisson processes (RPP) to model stochastic popularity dynamics and then incorporate it into the Bayesian framework for external factors inference and parameter estimation. PETM [30] improved RPP by introducing a power-law temporal relaxation function, an exponential reinforcement function and time mapping process. Gao et al. [48] split the complete diffusion process into many subprocesses and used RPP to model the subprocesses, which makes the proposed model efficient when trained on a single tweet. Mishra et. al. [49] present a hybrid predictor which combines Hawkes self-exciting point process for modeling each cascade and leverages feature-driven method for estimating the content virality, memory decay, and user influence. Later work HIP (Hawkes Intensity Process) [50] extended the original Hawkes process, which can explain the complex popularity history of each video according to its type of content, diffusion network, and sensitivity to promotion. As previously stated, the point process-based methods learn the intensity function for each event within the observation window, and learn the parameters by maximizing the occurrence probability, which can capture the dynamic process of the message re-sharing behavior, hence, have good comprehensibility. However, these methods hold the hypothesis that historical events will always excite future events, which is obviously not true in real life. Furthermore, these methods are not directly supervised by popularity, so there is a gap between modeling and prediction, which has hampered model performance in information cascade prediction.

Handcrafted feature-based methods: Theses methods extracted various handcrafted features from raw data. They typically include content features [22, 38, 51, 52], user features [25, 39, 53], structural features [27, 37, 54] and temporal features [28, 55], and then feed these features to discriminative machine learning algorithms to perform the cascades prediction tasks. Tsur et al. [38] demonstrated that combining content features with other types of features, e.g., temporal and structural features will reduce the prediction error. Bakshy et al. [25] studied the features related to early adopters and found that user features are informative predictors. Recently, in spite of exploring informative features, Shulman et al. [56] compared the predictive power of models using different sets of features and found that temporal features are the most predictive, almost as accurate as using content and user features. Cheng et al. [1] cast the information cascade size prediction as a classification task and concluded both temporal and structural features are almost

equally useful in predicting cascade size with an accuracy of 0.622 and 0.620, respectively. Summarizing, the performance of feature-based methods heavily depends on the hand-craft features, but there is not a standard and systematic way to design these features.

2.1.2 Deep learning-based methods

With the rapid advancement of deep learning in computer vision and natural language processing, researchers have developed a number of deep models to solve the problem of macro-level information cascades modeling and prediction. The key idea of such deep learning-based models is to automatically extract various diffusion features from the input cascades by leveraging different kinds of neural networks.

DeepCas [8] first demonstrated the effectiveness of deep neural networks in modeling information cascades. It first transformed the cascade graph as a set of node sequences by random walk and then automatically learned the structural features of individual graphs using GRU [40] and the attention mechanism. Li et al. extended DeepCas to DCGT [57] by incorporating content features. DeepHawkes [9] extracted temporal features by modeling diffusion paths via GRU rather than the random walks in DeepCas, and proposed the non-parametric time-decay effect to further improve the prediction performance, which bridged the gap between deep representation learning and the conventional Hawkes process. Gou et al. [58] proposed LSTMIC, which first converted the retweeting time series into several viewpoints, and then employed a long short-term memory (LSTM) architecture and pooling mechanism to extract sequential temporal features for information outbreak prediction. NT-GP [59] extracted node sequences from the user's activity log using the time decay sampling method, and then uses gated recurrent units (GRUs) to learn the temporal features from the sampling sequences and predict the target event's future diffusion range. The latest work TempCas [60] introduced a heuristic method for sampling full critical paths and it was shown to be more powerful than random walks and diffusion paths. It uses BiGRU with attention pooling for path embedding while modeling the short-term outbreaks and the impact of historical short-term outbreaks with an attention convolutional neural network (CNN) and an LSTM. Chen et al. [10] proposed the first graph neural networks (GNNs) based model called CasCN. It learned the structural and temporal information from sub-cascade graphs via a combination of graph convolutional network (GCN) [61] and LSTM, which also took into account the diffusion direction and time-decay effect. Later, some works [62, 63, 64] were built upon the CasCN through changing the graph kernel or using different sampling methods. For example, Cascade2Vec [62] improved the convolutional kernel of CasCN with the idea from graph Isomorphism network (GIN) [65] and residual networks [66]. Xu et al. proposed CasGCN [63], which first represented cascade graph as an in-coming graph and an out-coming graph, and then applied GCN to learn the structural features from both in-coming and out-coming

cascade graphs. The temporal features are learned through the normalization of diffusion time. CasSeqGCN [64] assumed that each sampled sub-cascade graph has the same topology but with a different state vector. Huang et al. proposed a graph sequence attention network – GSAN [67], which captured bi-directional and long dependencies between sub-cascade graphs via a collaboration of graph transformer block and a sequence transformer block. Another work [68]–CoupledGNN learned the cascading effect in information diffusion via coupled GNNs, towards capturing the interpersonal influence and individual user behavior based on the global graph. VaCas [69] first used the unsupervised graph wavelet to learn the structural information for cascade graphs, and employed variational autoencoder (VAE) [70] to enhance the cascade representation learning. And MUCas [42] tried to learn complete structural features from a multi-scale perspective.

Furthermore, some existing works attempt to extract temporal and structural information by performing both micro-level and macro-level tasks concurrently using multi-task learning [71] or reinforcement learning [72]. Also, some works have emerged to solve the general problem inherent in deep cascade learning, such as catastrophic forgetting [73], long-tail data distribution [74], and model generalization [75, 76].

2.2 Rumor Detection

The problem of rumor (or fake news/information, misinformation) detection is an important research topic in recent social media studies and receives increased attention in various disciplines including politics [18], finance [19], marketing [12], healthcare [77], etc. “Rumor" is usually defined as a misleading story or misinterpretation of information, circulating among communities and pertaining to an object, event, or issue in public concern [43].

2.2.1 Conventional methods

Handcrafted feature-based methods: Most of earlier works extracted various hand-crafted features from raw data, which can be typically summarized as two types: (1) content features extracted from both text (e.g. characters, words, sentences and documents) and visual elements (e.g. images and videos), which can be further partitioned as lexical features [23, 29, 78], syntactic features [23, 79, 80], topic features [81], visual statistical features [24, 77], and visual content features [82]; and (2) social context features extracted from the user behavior and the diffusion network, which reflect the relationship among users and describe the diffusion process of a rumor, including user features [23, 26, 83], propagation features [29, 81, 84], and temporal features [29, 85]. After feature engineering, the selected features are used in discriminative machine learning algorithms (e.g., random forest, naive Bayes, and support vector machines) to classify the news or tweets.

Rumors aim to arouse much attention and stimulate the public mood. Therefore, their texts/images/videos tend to have certain patterns in contrast to truth. Zhao et al. [78] discovered two types of language patterns in rumors, i.e., inquiry and correction patterns, and detected the patterns of rumor messages through supervised feature selection on a set of labeled messages. Wu et al. [81] defined a set of topic features to summarize semantics and trained a Latent Dirichlet Allocation (LDA) model for detecting rumors on Weibo. Towards a more comprehensive understanding of the text on social media, existing works also come up with textual features derived from social media platforms, apart from general textual features, such as source links [77] and emotions [23]. As for visual content features, Jin et al. [82] found that images in rumors and non-rumors are visually distinctive on their distributions and propose five visual features to measure the rumors, i.e., visual clarity score, visual coherence score, visual similarity distribution histogram, visual diversity score, and visual clustering score. Social context features are derived from the social connection characteristics of social media. Rumors are usually created by a few users and spread by a large number of users. Therefore, user profiles are commonly used to measure the user's characteristics and credibility. For example, Castillo et al. [23] first identified the credibility of tweets on Twitter based on user features. Diffusion patterns, i.e., structural patterns and temporal patterns, are also shown to be effective for detecting rumors. Kwon et al. [29] extended the work of [23] by proposing 15 structural features extracted from the diffusion network and the user friendship network. In the work [85], the authors proposed a method for discretizing time and capturing the variation of temporal features associated with rumors.

However, the performance of feature-based methods heavily depends on the handcraft features, which lacks a standard and systematic way to design general features across platforms and to deal with different types of rumors. In fact, the conclusions of existing works usually contradict each other, primarily due to the differences between different types of datasets. For example, Yang et al. [86] designed a set of features (e.g., client-based features and location-based features) based on Weibo, whose users are mainly restricted to China. It is therefore difficult to use these features for detecting rumors spread on Twitter and Facebook due to the differences in languages, clients' and users' geographic distributions, etc.

Credibility Propagation-based methods: Inspired by the work of truth discovery that aims to find truth with conflicting information, this line of methods consists of two main steps, i.e., (1) credibility network construction and (2) credibility propagation. The underlying assumption of these methods is that the credibility of news is highly related to the reliability of relevant social media posts, and both homogeneous and heterogeneous credibility networks can be built for the propagation process. Homogeneous credibility networks consist of a single type of entity, such as posts and events. In contrast, heterogeneous credibility networks involve different

types of entities, such as posts, sub-events, and events. Gupta et al. [24] first introduced a PageRank-like credibility propagation algorithm by encoding users' credibility and tweets' implications on a user-tweet-event information network. Inspired by the idea of linking entities altogether and leveraging inter-entity connections for credibility propagation, Jin et al. [31] proposed a three-layer hierarchical credibility network, which includes news aspects and utilizes a graph optimization framework to infer event credibility. The work in [32] found that relations between messages on microblogs (i.e. support and oppose) are crucial for evaluating the truthfulness of news events, and built a homogeneous credibility network among tweets to guide the process of credibility evaluation. While comparing with direct classification on the individual entity, credibility propagation-based methods may leverage the inter-entity relations for robust detection results. However, the performance of these methods strongly relies on the constructed credibility network.

2.2.2 Deep learning-based methods

The deep learning-based models have shown improved performance over traditional methods due to their enhanced ability to automatically representation learning. Most existing deep learning-based methods are content-aware that mainly focused on extracting textual features [33, 87, 88, 89, 90] and visual features [91, 92] from news content, user comments, and images, etc. Ma et al. [33] proposed the first deep learning-based rumor detection model, which applies recurrent neural networks (RNN) to model rumors as varied length time series aimed to learn both textual and temporal features from raw data and thus detect rumors. Shu et al. [87] proposed a co-attention network to exploit both news content and user comments for rumor detection while discovering explainable sentences. Jin et al. [91] presented a model to extract the visual, textual, and social context features, which are fused by the attention mechanism. Moreover, researchers also employed other deep learning techniques, such as multi-task learning [93], transformer [94, 95, 96, 97], and knowledge enhancement [98], to learn more robust content-aware features for rumor detection. However, rumors are intentionally written by mimicking real news [99], which makes content-aware methods hard to further improve detection performance due to the lack of necessary domain knowledge.

Recently, a few works have tried to exploit diffusion patterns in news spreading for rumor detection, e.g., temporal features [44, 100, 101] and structural features [16, 34, 44, 102]. For example, Liu et al. [100] presented a time series classifier with RNN and CNN to predict whether a given news story is fake at an early stage, taking common user characteristics and propagation paths into consideration. Song et al. [101] proposed a temporal propagation-based model that can distinguish rumors from true news through modeling dynamic evolution patterns of news. As for the structural features, Ma et al. [34] presented a tree-structured RNN to catch the hidden representations from both propagation structures and text contents. Inspired

by the success of graph neural networks in information cascades modeling [10, 71], Bian et al. [16] proposed a graph convolutional network [61]-based model that can learn global structural relationships of rumor dispersion. He et al. [103] improved the work [16] by using event augmentation and contrastive learning. Similarly, Lu et al. [102] improved the work of [100] by calculating the similarity between users and used a graph-aware attention network for rumor detection.

Furthermore, researchers began to consider both structural and temporal features for rumor detection. We [44] introduced a hierarchical diffusion modeling model by extracting both temporal features and propagation structures from the microscopic diffusion and macroscopic diffusion jointly. In addition, some researchers realized that users play significant roles in rumor spreading. For example, we proposed PLRD [45], which extracted social homophily, influence, and susceptibility of users from the user interaction network for rumor detection. UMLARD [46] improved PLRD by considering the disentangled feature learning and introducing textual features. Dou et al. [104] proposed a user preference-aware rumor detection model to learn user endogenous preference and exogenous context from users' historical posts and reply network, respectively.