Learning to Predict Opinion Share in Social Networks

Masahiro Kimura
Department of Electronics
and Informatics
Ryukoku University
Otsu 520-2194, Japan
kimura@rins.ryukoku.ac.jp

Kazumi Saito
School of Administration
and Informatics
University of Shizuoka
Shizuoka 422-8526, Japan
k-saito@u-shizuoka-ken.ac.jp

Kouzou Ohara
Department of Integrated
Information Technology
Aoyama Gakuin University
Kanagawa 229-8558, Japan
ohara@it.aoyama.ac.jp

Hiroshi Motoda
Institute of Scientific and
Industrial Research
Osaka University
Osaka 567-0047, Japan
motoda@ar.sanken.osaka-u.ac.jp

Abstract

We address the problem of predicting the expected opinion share over a social network at a target time from the opinion diffusion data under the value-weighted voter model with multiple opinions. The value update algorithm ensures that it converges to a correct solution and the share prediction results outperform a simple linear extrapolation approximation when the available data is limited. We further show in an extreme case of complete network that the opinion with the highest value eventually takes over, and the expected share prediction problem with uniform opinion value is not well-defined and any opinion can win.

Introduction

Blogosphere and sites such as for social networking, knowledge-sharing and media-sharing in the World Wide Web have enabled to form various kinds of large social networks, through which behaviors, ideas and opinions can spread. Thus, substantial attention has been directed to investigating the spread of influence in these networks (Leskovec, Adamic, and Huberman 2007; Crandall et al. 2008; Wu and Huberman 2008).

The representative problem is the *influence maximization problem*, that is, the problem of finding a limited number of influential nodes that are effective for the spread of information through the network and new algorithmic approaches have been proposed under different model assumptions, e.g., descriptive probabilistic interaction models (Domingos and Richardson 2001; Richardson and Domingos 2002), and basic diffusion models such as *independent cascade (IC) model* and the *linear threshold (LT) model* (Kempe, Kleinberg, and Tardos 2003; Kimura et al. 2010; Chen, Wang, and Yang 2009). This problem has good applications in sociology and "viral marketing" (Agarwal and Liu 2008). The models used above allow a node in the network to take only one of the two states, i.e., either active or inactive, because the focus is on *influence*.

However, application such as an on-line competitive service in which a user can choose one from multiple choices/decisions requires a model that handles multiple states. Further, it is important to consider the value of each

choice, e.g., quality, brand, authority, etc. because this impacts other's choice. We formulate this problem as a value-weighted K opinion diffusion problem and provides a way to accurately predict the expected share of the opinions at a future target time T (before an consensus is reached) from a limited amount of observed data.

A good model for opinion dynamics would be a voter model. It is one of the most basic stochastic process model, and has the same key property with the *linear threshold (LT) model* that a node decision is influenced by its neighbor's decision, i.e., a person changes its opinion by the opinions of its neighbors. In the basic voter model which is defined on an undirected network, each node initially holds one of K opinions, and adopts the opinion of a randomly chosen neighbor at each subsequent discrete time-step.

There has been a variety of work on the voter model. Dynamical properties of the basic model, including how the degree distribution and the network size affect the mean time to reach consensus, have been extensively studied (Liggett 1999; Sood and Redner 2005) from mathematical point of view. Several variants of the voter model are also investigated (Castellano, Munoz, and Pastor-Satorras 2009; Yang et al. 2009) and non equilibrium phase transition is analyzed from physics point of view. Yet another line of work extends the voter model and combine it with a network evolution model (Holme and Newman 2006; Crandall et al. 2008). The major interests there are different from what this paper intends to address, i.e., share prediction at a specific time T with opinion values considered.

Even-Dar and Shapira (2007) investigated the influence maximization problem (maximizing the spread of the opinion that supports a new technology) under the basic *voter model* with two ($K = 2$) opinions (one in favor of the new technology and the other against it) at a given target time T. They showed that the most natural heuristic solution, which picks the nodes in the network with the highest degree, is indeed the optimal solution, under the condition that all nodes have the same cost. This work is close to ours in that it measures the influence at a specific time T but is different in all others (no share prediction, no value considered, $K = 2$, no asynchronous update and no learning).

To the best of our knowledge, there has been no study that tried to predict the future opinion shares from the limited observed data in machine learning framework for the problem

of modeling the diffusion of several competitive opinions in a social network based on the voter model with opinion values considered. We learn the values of opinions from the limited amount of observed opinion diffusion data (i.e., data from 0 to T_0) and use the estimated values to predict the future (i.e., share at $T \ (> T_0)$). We show that the proposed approach works very satisfactorily using two real world social networks, and further a simple theoretical analysis reveals that it is indeed crucial to consider the opinion values and accurately estimate them for share prediction.

Our contribution is that 1) we proposed an algorithm that ensures the global optimal solution for the opinion value estimation from the observed opinion diffusion data, 2) we showed that the estimated model can accurately predict the future expected opinion share and outperforms the simple linear extrapolation prediction, and that, in the extreme case where all the nodes are connected to each other (i.e., complete network), 3) the opinion share prediction problem is not well-defined without introduction of opinion values and any opinion can prevail, and 4) the consensus is reached at which the opinion with the highest value wins and all the others die.

Opinion Dynamics

We consider the diffusion of opinions in a social network represented by an undirected (bidirectional) graph $G = (V, E)$ with self-loops. Here, V and $E \ (\subset V \times V)$ are the sets of all the nodes and links in the network, respectively. For a node $v \in V$, let $\Gamma(v)$ denote the set of neighbors of v in G, that is, $\Gamma(v) = \{u \in V; \ (u, v) \in E\}$. Note that $v \in \Gamma(v)$.

Voter Model

According to the work (Even-Dar and Shapria 2007), we recall the definition of the basic voter model with two opinions on network G. In the voter model, each node of G is endowed with two states; opinions 1 and 2. The opinions are initially assigned to all the nodes in G, and the evolution process unfolds in discrete time-steps $t = 1, 2, 3, \cdots$ as follows: At each time-step t, each node v picks a random neighbor u and adopts the opinion that u holds at time-step $t - 1$.

More formally, let $f_t : V \rightarrow \{1, 2\}$ denote the *opinion distribution* at time-step t, where $f_t(v)$ stands for the opinion of node v at time-step t. Then, $f_0 : V \rightarrow \{1, 2\}$ is the initial opinion distribution, and $f_t : V \rightarrow \{1, 2\}$ is inductively defined as follows: For any $v \in V$,

$$\begin{cases} f_t(v) = 1, & \text{with probability } \frac{n_1(t-1,v)}{n_1(t-1,v) + n_2(t-1,v)}, \\ f_t(v) = 2, & \text{with probability } \frac{n_2(t-1,v)}{n_1(t-1,v) + n_2(t-1,v)}, \end{cases}$$

where $n_k(t, v)$ is the number of v's neighbors that hold opinion k at time-step t for $k = 1, 2$.

Value-weighted Voter Model

We extend the original voter model for our purpose. In our model, the total number of opinions is set to $K \ (\geq 2)$, and each node of G is endowed with $(K + 1)$ states; opinions

$1, \cdots, K$, and *neutral* (i.e., no-opinion state). We consider that a node is *active* when it holds an opinion k, and a node is *inactive* when it does not have any opinion (i.e., its state is neutral). We assume that nodes never switch their states from active to inactive. In order to discuss the competitive diffusion of K opinions, we introduce the *value parameter* $w_k \ (> 0)$ for each opinion k. In the same way as the original voter model, let $f_t : V \rightarrow \{0, 1, 2, \cdots, K\}$ denote the opinion distribution at time t, where opinion 0 denotes the neutral state. We also denote by $n_k(t, v)$ the number of v's neighbors that hold opinion k at time t for $k = 1, 2, \cdots, K$, i.e.,

$$n_k(t, v) = |\{u \in \Gamma(v); \ f_t(u) = k\}|.$$

We start the evolution process from an initial state in which each opinion is assigned to only one node and all other nodes are in the neutral state. Given a target time T, the evolution process unfolds in the following way. In general, each node v considers changing its opinion based on the current opinions of its neighbors at its $(j-1)$th update-time $t_{j-1}(v)$, and actually changes its opinion at the jth update-time $t_j(v)$, where $t_{j-1}(v) < t_j(v) \leq T$, $j = 1, 2, 3, \cdots$, and $t_0(v) = 0$. It is noted that since node v is included in its neighbors by definition, its own opinion is also reflected. The jth update-time $t_j(v)$ is decided at time $t_{j-1}(v)$ according to the exponential distribution of parameter λ (we simply use $\lambda = 1$ for any $v \in V$ in our experiments)[1]. Then, node v changes its opinion at time $t_j(v)$ as follows: If node v has at least one active neighbor at time $t_{j-1}(v)$,

$$f_{t_j(v)}(v) = k, \quad \text{with probability } \frac{w_k \, n_k(t_{j-1}(v), v)}{\sum_{k'=1}^{K} w_{k'} \, n_{k'}(t_{j-1}(v), v)}$$

for $k = 1, \cdots, K$, otherwise,

$$f_{t_j(v)}(v) = 0, \quad \text{with probability } 1.$$

Note here that $f_t(v) = f_{t_{j-1}(v)}(v)$ for $t_{j-1}(v) \leq t < t_j(v)$. If the next update-time $t_j(v)$ pasts T, that is, $t_j(v) > T$, then the opinion evolution of v is over. The evolution process terminates when the opinion evolution of every node in G is over.

Opinion Share Prediction

Based on our opinion dynamics model, we investigate the problem of predicting how large a share each opinion will have at a future target time T when the opinion diffusion is observed from $t_0(= 0)$ to $T_0 \ (< T)$. Let $\mathcal{D}_{T_0}$ be the observed opinion diffusion data in time-interval $[0, T_0]$, that is,

$$\mathcal{D}_{T_0} = \{(v, t, f_t(v)); \ v \in V, \ t = 0, t_1(v), \cdots, t_{J_v}(v)\}.$$

Note that $t_{J_v}(v) \leq T_0$ for every $v \in V$. We define the *population* $h_k(t)$ of opinion k at time t by

$$h_k(t) = |\{v \in V; \ f_t(v) = k\}|$$

for $k = 1, 2, \cdots, K$.

---

[1]Note that this is equivalent to picking a node randomly and updating its opinion in turn $|V|$ times.
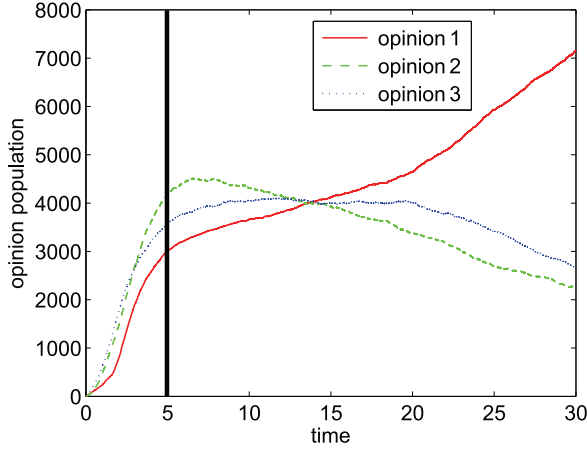
Figure 1: An example of opinion population curves in the blog network for $K = 3$.

Figure 1 shows an example of opinion population curves $h_1(t)$, $h_2(t)$, $h_3(t)$ for $K = 3$ in the blog network (see the section of "Experimental Evaluation" below), where $w_1 = 1.5$, $w_2 = 1.0$, $w_3 = 1.1$. Here, if we set $T_0 = 5$ and $T = 30$, we are able to observe $\mathcal{D}_5$ and thus $\{h_k(t); 0 \leq t \leq 5\}$ for $k = 1, 2, 3$ and the problem is to predict $h_1(30)$, $h_2(30)$, $h_3(30)$. Note that although the opinion dynamics is stochastic, we found that the variance of the value of $h_k(30)$ ($k = 1, 2, 3$) is relatively small for $T_0 = 5$. We can easily see from Figure 1 that the naive time-series analysis method does not work well for this prediction problem. Thus, it is crucial to accurately estimate the values of w_1, w_2 and w_3 from the observed opinion diffusion data.

We define the *share* $g_k(t)$ of opinion k at time t by

$$g_k(t) = \frac{h_k(t)}{\sum_{k'=1}^{K} h_{k'}(t)}.$$

Since our opinion dynamics model defines a stochastic process, we consider the problem of predicting the *expected share* of each opinion k at a given target time T, denoted by $\overline{g}_k(T)$. For solving this problem, we develop a method that effectively estimates the values of *value parameters* $w_1, \cdots, w_K$ from the observed data $\mathcal{D}_{T_0}$.

Simple Case Analysis

We analyze the effects of value parameters at the time t where all nodes have become active for an extreme case in which the network is complete, i.e., neighbors of each node cover the whole network. According to the previous work (e.g., (Sood and Redner 2005)), the expected share change $dg_k(t)$ can be calculated as follows:

$$
dg_k(t) = \frac{1}{|V|}(1 - g_k(t))\frac{g_k(t)w_k}{\sum_{k'=1}^{K} g_{k'}(t)w_{k'}}
$$
$$
- \frac{1}{|V|}g_k(t)\left(1 - \frac{g_k(t)w_k}{\sum_{k'=1}^{K} g_{k'}(t)w_{k'}}\right)
$$

$$
= \frac{1}{|V|}\left(\frac{g_k(t)w_k}{\sum_{k'=1}^{K} g_{k'}(t)w_{k'}} - g_k(t)\right). \quad (1)
$$

Now, let k^* be the opinion with the highest value parameter such that $w_{k^*} > w_k$ for all the other opinion k ($k \neq k^*$). Then, we can obtain the following inequality from Eq. (1) when $g_k(t) > 0$ for all k:

$$
dg_{k^*}(t) = \frac{g_{k^*}(t)w_{k^*}}{|V|\sum_{k=1}^{K} g_k(t)w_k}\left(1 - \sum_{k=1}^{K} g_k(t)\frac{w_k}{w_{k^*}}\right)
$$
$$
> \frac{g_{k^*}(t)w_{k^*}}{|V|\sum_{k=1}^{K} g_k(t)w_k}\left(1 - \sum_{k=1}^{K} g_k(t)\right) = 0.
$$

Here note that $w_k/w_{k^*} < 1$ for $k \neq k^*$. Therefore, unless $g_{k^*}(t) = 0$, the opinion k^* is expected to finally prevail the others, regardless of its current share since the function $g_{k^*}(t)$ is expected to increase as time passes until each of the other opinion shares becomes 0. This result suggests that it is crucially important to accurately estimate the value parameter of each opinion from the observed data $\mathcal{D}_{T_0}$. Moreover, we can see that if the value parameters are uniform, any opinion can become a winner. These observations imply that the expected share prediction problem can be well-defined only when the opinion values are non-uniform. We conjecture that results will be similar for more realistic networks, although the above analysis is valid for a complete network.

Consensus Time Analysis

We further analyze the consensus time by using the above simple case. For simplicity, we assume that $w_k = w$ if $k \neq k^*$, i.e., the values of the other value parameters are the same. Let r be the ratio of the value parameters defined by $r = w/w_{k^*}$; then, by regarding $1/|V|$ as a time step dt (e.g., (Sood and Redner 2005)), we can obtain the following differential equation for $g_{k^*}(t)$ from Eq. (1):

$$
\frac{dg_{k^*}(t)}{dt} = \frac{g_{k^*}(t)}{r(1 - g_{k^*}(t)) + g_{k^*}(t)} - g_{k^*}(t)
$$
$$
= \frac{(1 - r)g_{k^*}(t)(1 - g_{k^*}(t))}{r + (1 - r)g_{k^*}(t)}.
$$

From this differential equation, we can easily derive the following solution:

$$
\frac{r}{1-r}\log(g_{k^*}(t)) - \frac{1}{1-r}\log(1 - g_{k^*}(t)) = t + C,
$$

where C stands for a constant of integration. Figure 2 shows examples of expected share curves based on the above solution with different ratios of the value parameters, where the ratio r is set to $r = 1 - 2^{-a}$ ($a = 1, 2, 3, 4, 5$), and each curve is plotted from $t = 0$ by assuming $g_{k^*}(0) = 0.01$ until $t = T$ that satisfies $g_{k^*}(T) = 0.99$. From this figure, we can see that the consensus time is quite short when the ratio r is small, while it takes somewhat longer when the ratio r approaches to 1. More importantly, this result indicates that the consensus time of our model is extremely short even
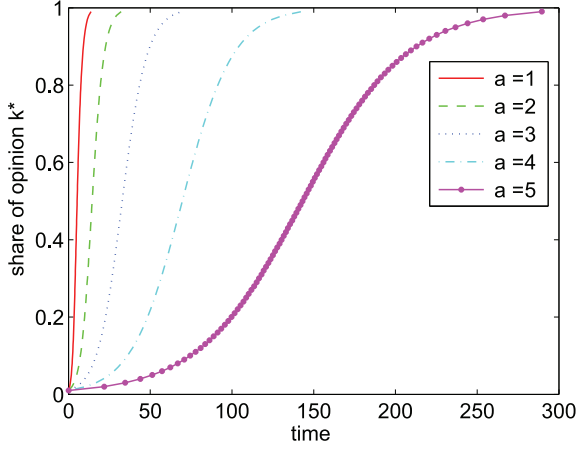
Figure 2: Examples of expected share curves.

when the ratio r is close to 1, compared with the basic voter model studied in previous work (e.g., (Even-Dar and Shapria 2007)). Therefore, we consider that the voter model can be more practical by introducing the value parameters.

Learning Method

For a given observed opinion diffusion data $\mathcal{D}_{T_0}$, we focus on the competitive opinion diffusion data $\mathcal{C}_{T_0}$ defined by

$$\mathcal{C}_{T_0} = \{(v, t, f_t(v)) \in \mathcal{D}_{T_0}; |\{u \in \Gamma(v); \ f_t(u) \neq 0\}| \geq 2\}.$$

Then, from the evolution process of our model described in the previous section, we can obtain the following likelihood function[2]:

$$\mathcal{L}(\boldsymbol{w}; \mathcal{C}_{T_0}) = \log \prod_{(v,t,k)\in\mathcal{C}_{T_0}} \frac{n_k(t,v)w_k}{\sum_{k'=1}^{K} n_{k'}(t,v)w_{k'}}, \quad (2)$$

where $\boldsymbol{w}$ stands for the K-dimensional vector of value parameters, i.e., $\boldsymbol{w} = (w_1, \cdots, w_K)$. Thus our estimation problem[3] is formulated as a maximization problem of the objective function $\mathcal{L}(\boldsymbol{w}; \mathcal{C}_{T_0})$ with respect to $\boldsymbol{w}$.

Note that the objective function $\mathcal{L}(\boldsymbol{w}; \mathcal{C}_{T_0})$ is invariant to positive scaling of the value parameter vector $\boldsymbol{w}$, and each value parameter w_k must be positive, as noted earlier. In order to formulate our maximization problem as an unconstrained optimization problem, we reparameterize each value parameter w_k by using a $(K-1)$-dimensional vector $\boldsymbol{z} = (z_1, \cdots, z_{K-1})$ as follows:

$$w_k = \begin{cases} \exp(z_k) & \text{if } k < K, \\ 1 & \text{if } k = K. \end{cases} \quad (3)$$

Namely, our estimation problem is formulated as an optimization problem of the objective function $\mathcal{L}_1(\boldsymbol{z}; \mathcal{C}_{T_0})$ ($= \mathcal{L}(\boldsymbol{w}; \mathcal{C}_{T_0})$) with respect to $\boldsymbol{z}$.

---

[2]Introduction of $\mathcal{C}_{T_0}$ is simply to avoid $\log(0/0)$ and $\log 1$.

[3]The delay time parameter λ can also be a parameter, but it can simply be estimated by averaging the time intervals for each node, and thus excluded from the estimation problem. Estimating this parameter is not critical to the current problem because its value simply contributes to scaling the time unit.

In order to derive our learning algorithm, we consider the following probability that the node v adopts the opinion k ($k < K$) at time t.

$$q_k(t, v) = \frac{n_k(t, v)\exp(z_k)}{n_K(t, v) + \sum_{k'=1}^{K-1} n_{k'}(t, v)\exp(z_{k'})} \quad (4)$$

Then, we can obtain the first-order derivative (gradient vector element) of $\mathcal{L}_1(\boldsymbol{z}; \mathcal{C}_{T_0})$ with respect to z_i as follows:

$$\frac{\partial \mathcal{L}_1(\boldsymbol{z}; \mathcal{C}_{T_0})}{\partial z_i} = \sum_{(v,t,k)\in\mathcal{C}_{T_0}} (\delta_{k,i} - q_i(t, v)),$$

where $\delta_{k,i}$ is the Kronecker's delta. Similarly, we can obtain the second-order derivative (Hessian matrix element) with respect to z_i and z_j as follows:

$$\frac{\partial^2 \mathcal{L}_1(\boldsymbol{z}; \mathcal{C}_{T_0})}{\partial z_i \partial z_j} = \sum_{(t,v,k)\in\mathcal{C}_{T_0}} (q_i(t, v)q_j(t, v) - \delta_{i,j}q_i(t, v)).$$

Here note that the following quadratic form of the Hessian matrix is non-positive for an arbitrary $(K-1)$-dimensional non-zero vector $\boldsymbol{x} = (x_1, \cdots, x_{K-1})$,

$$\sum_{i,j=1}^{K-1} \frac{\partial^2 \mathcal{L}_1(\boldsymbol{z}; \mathcal{C}_{T_0})}{\partial z_i \partial z_j} x_i x_j$$

$$= \sum_{(v,t,k)\in\mathcal{C}_{T_0}} \left(\left(\sum_{i=1}^{K-1} q_i(t,v)x_i \right)^2 - \sum_{i=1}^{K-1} q_i(t,v)x_i^2 \right)$$

$$= -\sum_{(v,t,k)\in\mathcal{C}_{T_0}} \sum_{i=1}^{K-1} q_i(t,v) \left(x_i - \sum_{j=1}^{K-1} q_j(t,v)x_j \right)^2$$

$$- \sum_{(v,t,k)\in\mathcal{C}_{T_0}} \left(1 - \sum_{i=1}^{K-1} q_i(t,v) \right) \left(\sum_{j=1}^{K-1} q_j(t,v)x_j \right)^2$$

$$\leq 0.$$

Thus we can guarantee that the solution of our problem is global optimal. Our implemetation employes a sandard Newton method. The algorithm of the proposed method is summarized below.

1. Initialize parameter vector $\boldsymbol{z}$ as $z_k = 0$ for $k = 1, \cdots, K-1$.

2. Calculate the gradient vector at the current parameter vector $\boldsymbol{z}$.

3. If the gradient vector is sufficiently small, i.e., $\sum_i (\partial \mathcal{L}_1(\boldsymbol{z}; \mathcal{C}_{T_0})/\partial z_i)^2 < \eta$, output the value parameters by using Eq. (3) then terminate. Otherwise, go to 4.

4. Calculate the Hessian matrix and its inverted matrix, and update the parameter vector $\boldsymbol{z}$ by multiplying the inverted matrix and the gradient vector, and return to 2.

Here η is a parameter for the termination condition. In our experiments, η is set to a sufficiently small number, i.e., $\eta = 10^{-12}$.
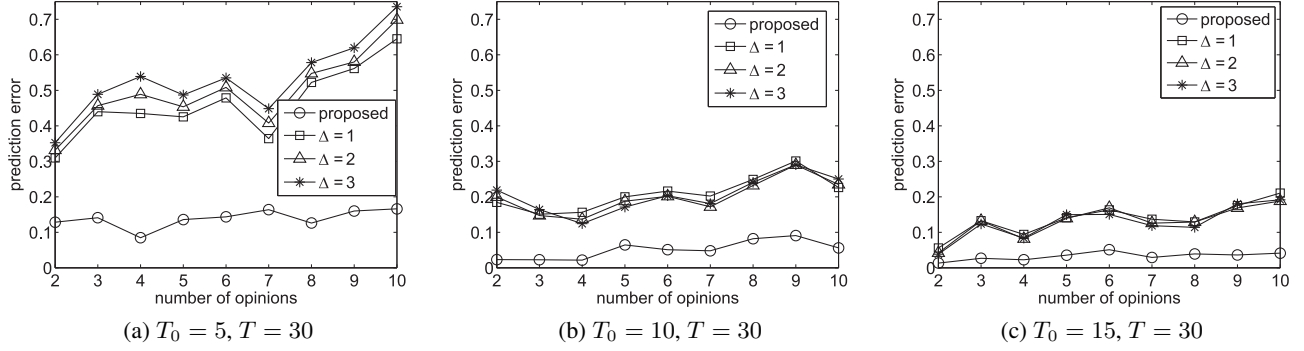
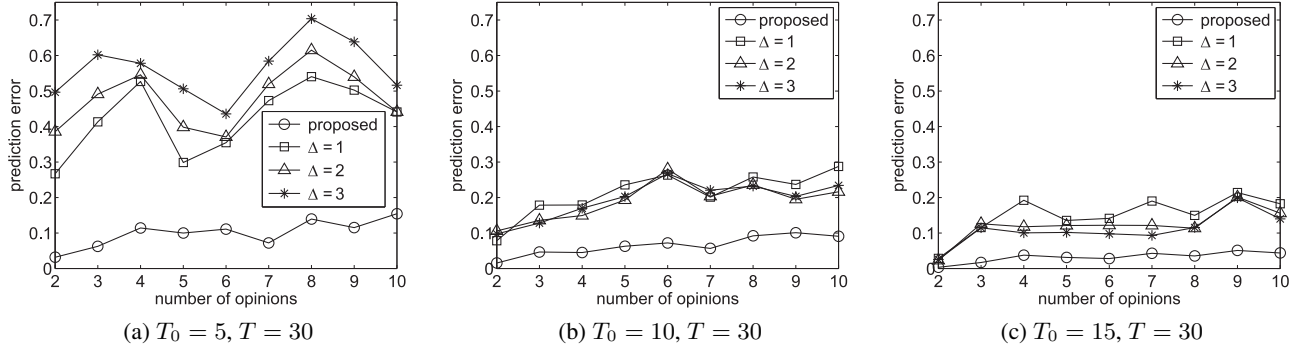Figure 3: Results for share prediction in the blog network.

Figure 4: Results for share prediction in the Wikipedia network.

Experimental Evaluation

Network Datasets and Experimental Settings

We employed two datasets of large real networks used in (Kimura, Saito, and Motoda 2009), which are bidirectional connected networks and exhibit many of the key features of social networks. The first one is a trackback network of Japanese blogs and had $12,047$ nodes and $79,920$ directed links (the blog network). The second one is a network of people that was derived from the "list of people" within Japanese Wikipedia, and had $9,481$ nodes and $245,044$ directed links (the Wikipedia network).

We varied $K = 2, 3, \cdots, 10$, and for each of them we predicted the expected share $\overline{g}_k(T)$ of opinion k ($k = 1, 2, \cdots, K$) for the observed data $\mathcal{D}_{T_0}$. We set $T = 30$, investigated the cases $T_0 = 5, 10, 15$, and selected the true value of each value parameter w_k from the interval $[0.5, 1.5]$ uniformly at random. We chose the top K nodes with respect to node degree ranking as the initial K nodes, and generated $\mathcal{D}_{T_0}$ by simulating the true model. After we have estimated the value of each w_k, we predicted the value of $\overline{g}_k(T)$ by simulating the model M times from $\mathcal{D}_{T_0}$ and taking their average, where we used $M = 100$. In fact, our preliminary experiments indicate that the result for $M = 100$ are not much different from those for $M = 1,000$ and $10,000$ in the blog and the Wikipedia networks. Note that the number of opinion updates amounts to tens of thousands for one

instance of $\mathcal{D}_{T_0}$, and thus no overfitting problem arises.

Comparison Methods and Evaluation Measure

Given the observed data $\mathcal{D}_{T_0}$, we can simply apply a linear extrapolation for predicting the expected share of opinion k at a target time T, since we can naively speculate that the recent trend for each opinion continues. Thus, we consider predicting the values of $\overline{g}_1(T), \cdots, \overline{g}_K(T)$, by estimating the value of the population $h_k(T)$ of opinion k at time T based on the linear extrapolation from the values of $h_k(T_0 - \Delta)$ and $h_k(T_0)$ for each k, where Δ is the parameter with $0 < \Delta \leq T_0$. We refer to this prediction method as the *naive linear method*. We evaluated the effectiveness of the proposed share prediction method by comparing it with the *naive linear method*.

Let $\widehat{g}_k(T)$ be the estimate of $\overline{g}_k(T)$ by a share prediction method. We measured the performance of the share prediction method by the prediction error $\mathcal{E}$ defined by

$$\mathcal{E} = \sum_{k=1}^{K} |\widehat{g}_k(T) - \overline{g}_k(T)|.$$

Experimental Results

Figures 3a, 3b, and 3c are the results for the blog network, and Figures 4a, 4b, and 4c for the Wikipedia network, where circles indicate the prediction errors of the proposed method,

and squares, triangles, and asterisks indicate the prediction errors of the *naive linear method* adopting $\Delta = 1$, $\Delta = 2$, and $\Delta = 3$, respectively. We conducted 10 trials varying the true values of value parameters for each K, and plotted the average of $\mathcal{E}$ over the 10 trials.

From these figures, we can see that the prediction error decreases as the observation time T_0 becomes longer and that the proposed method outperforms the *naive linear method* in every case. When $T_0 = 5$, the average prediction error of the proposed method was 0.139 for the blog network and 0.100 for the Wikipedia network, while that of the *naive method* was at least 0.465 and 0.424, respectively in case of $\Delta = 1$. When $T_0 = 15$, the average prediction error of the proposed method was 0.033 for the blog network and 0.032 for the Wikipedia network, while that of the *naive method* was at least 0.128 for the blog network and 0.110 for the Wikipedia network in case of $\Delta = 3$, which is comparable to those of the proposed method for $T_0 = 5$. Moreover, we observed that the proposed method accurately predicted the share at T even in the case that the share ranking at T_0 got reversed at the target time T as shown in Figure 1. This is attributed to the use of the estimated value parameters which take different values for different opinions, and is consistent with the aforementioned analysis on a complete network.

During the experiments we noticed that the time needed to reach the consensus gets longer when the difference between the largest and the second largest values of the value parameters is small. This can also be predicted by the consensus time analysis, i.e., considering the case where the highest two values are the same and the rest are also the same.

Consequently, we confirmed that the results of our theoretical analyses hold in real networks and that the proposed method outperforms the *naive linear method*. On average, the prediction error of the proposed method was about four times less for a given T_0. Besides, it achieved a comparable prediction accuracy in three times less observation time compared with the *naive linear method*.

Conclusion

We addressed the problem of how different opinions with different values spread over a social network and how their share changes over time in a machine learning setting using a variant of *voter model*, the value-weighted voter model with multiple opinions. The task is first to estimate the opinion values from the limited amount of observed data and the goal is to predict the expected opinion share at a future target time. We derived an algorithm that guarantees the global optimal solution for the opinion value estimation and showed using two real world social networks that the values are learnable from a small amount of observed data and the share prediction with use of the estimated values is satisfactorily accurate and outperforms the prediction by a simple linear extrapolation. Theoretical analysis for an extreme case where all the nodes are connected to each other (a complete network) revealed that the expected share prediction problem is well-defined only when the opinion values are non-uniform in which case the final consensus is winners-takes-all, i.e., the opinion with the highest value wins and all the others die, and when they are uniform, any opinion

can be a winner. Our immediate future work is to validate the credibility of the *voter model* using available real opinion propagation data.

References

Agarwal, N., and Liu, H. 2008. Blogosphere: Research issues, tools, and applications. *SIGKDD Explorations* 10:18–31.

Castellano, C.; Munoz, M. A.; and Pastor-Satorras, R. 2009. Nonlinear q-voter model. *Physical Review E* 80:041129.

Chen, W.; Wang, Y.; and Yang, S. 2009. Efficient influence maximization in social networks. In *Proceedings of KDD 2009*, 199–208.

Crandall, D.; Cosley, D.; Huttenlocner, D.; Kleinberg, J.; and Suri, S. 2008. Feedback effects between similarity and sociai infuence in online communities. In *Proceedings of KDD 2008*, 160–168.

Domingos, P., and Richardson, M. 2001. Mining the network value of customers. In *Proceedings of KDD 2001*, 57–66.

Even-Dar, E., and Shapria, A. 2007. A note on maximizing the spread of influence in social networks. In *Proceedings of WINE 2007*, 281–286.

Holme, P., and Newman, M. E. J. 2006. Nonequilibrium phase transition in the coevolution of networks and opinions. *Physical Review E* 74:056108.

Kempe, D.; Kleinberg, J.; and Tardos, E. 2003. Maximizing the spread of influence through a social network. In *Proceedings of KDD 2003*, 137–146.

Kimura, M.; Saito, K.; Nakano, R.; and Motoda, H. 2010. Extracting influential nodes on a social network for information diffusion. *Data Mining and Knowledge Discovery* 20:70–97.

Kimura, M.; Saito, K.; and Motoda, H. 2009. Blocking links to minimize contamination spread in a social network. *ACM Transactions on Knowledge Discovery from Data* 3:9.

Leskovec, J.; Adamic, L. A.; and Huberman, B. A. 2007. The dynamics of viral marketing. *ACM Transactions on the Web* 1:5.

Liggett, T. M. 1999. *Stochastic interacting systems: contact, voter, and exclusion processes*. New York: Spriger.

Richardson, M., and Domingos, P. 2002. Mining knowledge-sharing sites for viral marketing. In *Proceedings of KDD 2002*, 61–70.

Sood, V., and Redner, S. 2005. Voter model on heterogeneous graphs. *Physical Review Letters* 94:178701.

Wu, F., and Huberman, B. A. 2008. How public opinion forms. In *Proceedings of WINE 2008*, 334–341.

Yang, H.; Wu, Z.; Zhou, C.; Zhou, T.; and Wang, B. 2009. Effects of social diversity on the emergence of global consensus in opinion dynamics. *Physical Review E* 80:046108.