

Opinion Formation by Voter Model with Temporal Decay Dynamics

Masahiro Kimura¹, Kazumi Saito², Kouzou Ohara³, and Hiroshi Motoda⁴

¹ Department of Electronics and Informatics, Ryukoku University
Otsu 520-2194, Japan
kimura@rins.ryukoku.ac.jp

² School of Administration and Informatics, University of Shizuoka
Shizuoka 422-8526, Japan
k-saito@u-shizuoka-ken.ac.jp

³ Department of Integrated Information Technology, Aoyama Gakuin University
Kanagawa 252-5258, Japan
ohara@it.aoyama.ac.jp

⁴ Institute of Scientific and Industrial Research, Osaka University
Osaka 567-0047, Japan
motoda@ar.sanken.osaka-u.ac.jp

Abstract. Social networks play an important role for spreading information and forming opinions. A variety of voter models have been defined that help analyze how people make decisions based on their neighbors' decisions. In these studies, common practice has been to use the latest decisions in opinion formation process. However, people may decide their opinions by taking account not only of their neighbors' latest opinions, but also of their neighbors' past opinions. To incorporate this effect, we enhance the original voter model and define the temporal decay voter (TDV) model incorporating a temporary decay function with parameters, and propose an efficient method of learning these parameters from the observed opinion diffusion data. We further propose an efficient method of selecting the most appropriate decay function from among the candidate functions each with the optimized parameter values. We adopt three functions as the typical candidates: the exponential decay, the power-law decay, and no decay, and evaluate the proposed method (parameter learning and model selection) through extensive experiments. We, first, experimentally demonstrate, by using synthetic data, the effectiveness of the proposed method, and then we analyze the real opinion diffusion data from a Japanese word-of-mouth communication site for cosmetics using three decay functions above, and show that most opinions conform to the TDV model of the power-law decay function.

1 Introduction

Social networking services (SNSs) on the Internet, such as Facebook, Twitter and Digg, have become so popular and use of these services is now a part of our daily activities. Large networks formed by these services play an important role as a medium for spreading diverse information including news, ideas, opinions, and rumors [18,17,8,6]. Users of these services can share their interests or opinions to each other. The resulting social

networks and the information propagated therein have great influence on and drastically change our decision making processes and behaviors in daily life. Thus, many attempts have been made to investigate the spread of influence in social networks [15,5,21].

One such typical and well studied problem in social network analysis is the *influence maximization problem*, which is finding a limited number of influential nodes that are effective for spreading information [10,11,16,3,4]. What is common to these studies is that models used 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, we need a model in which a node can take multiple states for such applications in which a user can choose one from multiple choices. For example, a mobile phone user may change his/her current carrier to the one which the majority of his/her neighbors are using. To model this kind of opinion formation dynamics, a node in the network has to be able to take one of many possible choices as its state. A *voter model* would be the one which is most suitable for this purpose. It is one of the most basic stochastic process models, where a node decision is influenced by its neighbors' decisions [20,9,7,2,22]. We proposed two variants of voter model in our past work: the *value-weighted voter model* that considers opinion values [12], and the *value-weighted mixture voter model* that, in addition to the opinion values, considers the effect of anti-majoritarians, i.e., those people who do not agree with the majority and support the minority opinion [13].

In this paper we also address the problem of opinion formation on the social network, but we especially focus on the fact that our decision may be influenced not only by our neighbors' and our own latest opinions, but also by the neighbors' and our own past opinions. For example, assume that you and your friends have long supported a certain political party, but many of your friends have started changing their supporting party to a different one very recently. Under this situation, you may still stick to your opinion and keep supporting the party, or you may change your mind and follow your neighbors' opinions. This means that your current opinions are influenced not only by the neighbors' latest opinions but also by their past opinions including your own opinions. It is, thus, important to consider all the past opinions in making the current decision. Nonetheless, all the voter models including the two variants mentioned above consider only the latest opinions of its neighbors including itself when updating the opinion of a node.

With this in mind we enhance the original voter model and define the *temporal decay voter (TDV) model* that takes into account all the past opinions discounting the effect of older opinions by using a temporal decay function. The work most closely related to our approach would be the work by Koren [14] which is in the context of recommender systems, where several time drifting user preference models are proposed, some of which adopt a temporal decay function that discounts the effect of older ratings to items. The approach in Koren's work is, unlike our approach, cannot utilize all the past ratings given by a user for an identical item because the user-item matrix that they use does not allow multiple ratings to be stored. In addition, due to the framework of collaborative filtering, it requires the rating history involving multiple items, while our approach can model the temporal dynamics of opinions for a single item. Thus, it does not make sense to compare Koren's approach with ours.

Our major contribution is the following four: 1) the TDV model, 2) an algorithm of learning the parameters of the temporal decay function from the observed opinion spreading data, 3) a model selection method that determines the most appropriate decay function for given data, and 4) new finding regarding the decay model from the analysis of the real data. The model parameters are learned by an efficient iterative algorithm which maximizes the likelihood function. Three representative decay functions are employed, although the framework is not necessarily limited to them: the exponential decay, the power-law decay, and no decay. Which function, each with the optimized parameter values, is most appropriate for given data is determined based on the log likelihood ratio statistic. We evaluate the parameter learning and the model selection methods through extensive experiments using synthetic data with two TDV models, one with the exponential decay and the other with power-low decay. We then applied the methods to the real opinion spreading data from a Japanese word-of-mouth communication site for cosmetics using aforementioned three decay functions, and show that most opinions conform to the TDV model of the power-law decay function.

The paper is organized as follows. We define the TDV model in Section 2 and explain how the model parameters are learned and the most appropriate model is selected in Section 3. The performance of parameter learning and model selection using the synthetic data is reported in Section 4 and the finding from the analysis of real data is reported in Section 5. We end this paper by summarizing the main result in Section 6.

2 Voter Model with Temporal Decay Dynamics

We define the TDV (Temporal Decay Voter) model. Let $G = (V, E)$ be a directed network with self-loops, where V and $E (\subset V \times V)$ are the sets of all nodes and links in the network, respectively. Here, $(u, v) \in E$ denotes a (directed) link from node u to node v . When there is a link (u, v) , we assume that v can be influenced by its neighbor u in opinion formation process. For a node $v \in V$, let $B(v)$ denote the set of neighbors of v in G , that is,

$$B(v) = \{u \in V; (u, v) \in E\}.$$

Note that $v \in B(v)$. Given an integer K with $K \geq 2$, we consider the spread of K opinions (opinion 1, \dots , opinion K) on G , where each node holds exactly one of the K opinions at any time $t (\geq 0)$. We assume that each node of G initially holds one of the K opinions with equal probability at time $t = 0$. We denote by

$$g_t : V \rightarrow \{1, \dots, K\}$$

the *opinion distribution* at time t , where $g_t(v)$ stands for the opinion of node v at time t . Note that g_0 stands for the initial opinion distribution. For any $v \in V$ and $k \in \{1, 2, \dots, K\}$, let $U_k(t, v)$ be the set of v 's neighbors that hold opinion k as its latest opinion (before time t), i.e.,

$$U_k(t, v) = \{u \in B(v); \varphi_t(u) = k\},$$

where $\varphi_t(u)$ is the latest opinion of u (before time t).

2.1 Voter Model

We first recall the definition of the voter model (see, e.g., [13]), which is one of the standard models of opinion dynamics, where K is usually set to 2. The evolution process of the voter model is defined as follows:

1. At time 0, each node v independently decides its update time t according to some probability distribution such as an exponential distribution with parameter $\gamma_v = 1$.¹ The successive update time is determined similarly at each update time t .
2. At an update time t , the node v adopts the opinion of a randomly chosen neighbor u , i.e.,

$$g_t(v) = \varphi_t(u).$$

3. The process is repeated from the initial time $t = 0$ until the next update-time passes a given final-time T_1 .

We note that in the voter model each individual tends to adopt the majority opinion among its neighbors.² Here note that the definition of one’s neighbors include oneself because of the existence of self loop. Thus, we can extend the original voter model with 2 opinions to a voter model with K opinions by replacing Step 2 with: At an update time t , the node v selects one of the K opinions according to the probability distribution,

$$P(g_t(v) = k) = \frac{|U_k(t, v)|}{|B(v)|}, \quad (k = 1, \dots, K). \tag{1}$$

2.2 Temporal Decay Voter Model

As mentioned earlier, people may decide their opinions by taking account not only of their neighbors’ latest opinions, but also of their neighbors’ past opinions including their own opinions. In order to model this kind of situation, for any $t > 0$ and $v \in V$, we consider the set $M(t, v)$ consisting of the time $\tau (< t)$ at which an individual (a node) v manifested his/her opinion. For $k = 1, \dots, K$, we also consider a subset of $M(t, v)$,

$$M_k(t, v) = \{\tau \in M(t, v); g_\tau(v) = k\},$$

where $M_k(t, v)$ is the set of node v ’s opinion manifestation time instances before time t in which v takes opinion k . Now, we can define a voter model which takes all the past opinions into consideration. In this model, Eq. (1) is replaced with

$$P(g_t(v) = k) = \frac{1 + \sum_{u \in B(v)} |M_k(t, u)|}{K + \sum_{u \in B(v)} |M(t, u)|}, \quad (k = 1, \dots, K), \tag{2}$$

where we employed a Bayesian prior known as the Laplace smoothing. Here we note that the Laplace smoothing of Eq. (2) corresponds to the assumption that each node initially holds one of the K opinions with equal probability at time $t = 0$. Note also that the

¹ This assumes that the average delay time is 1.

² In reality there may be a case that one changes its opinion to a medium one (say 3) listening to two opposite opininons (say 1 and 5). The voter model does not consider this possibility unless at least one of the neighbors has already the medium opinion (3).

Laplace smoothing corresponds to a special case of Dirichlet distributions that are very often used as prior distributions in Bayesian statistics, and in fact the Dirichlet distribution is the conjugate prior of the categorical distribution and multinomial distribution. We refer to this voter model as the *base TDV model*.

Thus far, we assumed that all the past opinions are equally weighted. However, it is naturally conceivable that the quite old opinions have almost no influence. Older opinions are less influential in general. In order to reflect this kind of effects into the model, we consider introducing some decay functions. The simplest one is an exponential decay function defined by

$$\rho(\Delta t; \lambda) = \exp(-\lambda \Delta t), \tag{3}$$

where $\lambda \geq 0$ is a parameter and $\Delta t = t - \tau$ stands for the time difference between the opinion adoption time t and the opinion manifestation time τ . Another natural one would be a power-law decay function defined by

$$\rho(\Delta t; \lambda) = (\Delta t)^{-\lambda} = \exp(-\lambda \log \Delta t), \tag{4}$$

where $\lambda \geq 0$ is a parameter.

Now, we construct a more general decay function. For a given positive integer J , let $f_1(\Delta t), \dots, f_J(\Delta t)$ be functions on $(0, +\infty)$ such that $1, f_1(\Delta t), \dots, f_J(\Delta t)$ are linearly independent, that is, if $\lambda_0, \lambda_1, \dots, \lambda_J$ are real numbers and satisfy

$$\lambda_0 + \sum_{j=1}^J \lambda_j f_j(\Delta t) = 0, \quad (\forall \Delta t \in (0, +\infty)),$$

then $\lambda_0 = \lambda_1 = \dots = \lambda_J = 0$. We then consider a J -dimensional feature vector,

$$\mathbf{F}_J(\Delta t) = (f_1(\Delta t), \dots, f_J(\Delta t))^T,$$

where \mathbf{a}^T denote the transpose of column vector \mathbf{a} . For a J -dimensional real column vector with non-negative elements,

$$\boldsymbol{\lambda}_J = (\lambda_1, \dots, \lambda_J)^T,$$

which is a parameter vector, we define a decay function $\rho(\Delta t; \boldsymbol{\lambda}_J)$ by

$$\rho(\Delta t; \boldsymbol{\lambda}_J) = \exp\left(-\boldsymbol{\lambda}_J^T \mathbf{F}_J(\Delta t)\right), \tag{5}$$

where the matrix operations are used. Representative candidates of feature vector $\mathbf{F}_J(\Delta t)$ include

$$\mathbf{F}_1(\Delta t) = \Delta t, \quad \mathbf{F}_1(\Delta t) = \log \Delta t, \quad \mathbf{F}_1(\Delta t) = (\Delta t)^2$$

for $J = 1$,

$$\mathbf{F}_2(\Delta t) = (\Delta t, \log \Delta t)^T, \quad \mathbf{F}_2(\Delta t) = (\Delta t, (\Delta t)^2)^T, \quad \mathbf{F}_2(\Delta t) = (\log \Delta t, (\Delta t)^2)^T$$

for $J = 2$,

$$\mathbf{F}_3(\Delta t) = (\Delta t, \log \Delta t, (\Delta t)^2)^T$$

for $J = 3$, etc. Note that $\rho(\Delta t; \lambda_J)$ becomes the exponential decay function if $J = 1$ and $F_J(\Delta t) = \Delta t$, and the power-law decay function if $J = 1$ and $F_J(\Delta t) = \log \Delta t$.

Using our general decay function $\rho(\Delta t; \lambda_J)$ (see Eq. (5)), we define the TDV (Temporal Decay Voter) model in the following way. In this model, Eq. (1) is replaced with

$$P(g_i(v) = k) = \frac{1 + \sum_{u \in B(v)} \sum_{\tau \in M_k(t,u)} \rho(t - \tau; \lambda_J)}{K + \sum_{u \in B(v)} \sum_{\tau \in M(t,u)} \rho(t - \tau; \lambda_J)}, \quad (k = 1, \dots, K). \quad (6)$$

Here note that Eq. (6) is reduced to Eq. (2) when λ_J is the J -dimensional zero-vector $\mathbf{0}_J$, that is, the TDV model of $\lambda_J = \mathbf{0}_J$ coincides with the base TDV model.

3 Learning Method

We consider the problem of identifying the TDV model on network G from an observed data \mathcal{D}_{T_0} in time-span $[0, T_0]$, where \mathcal{D}_{T_0} consists of a sequence of (k, t, v) such that node v changed its opinion to opinion k at time t for $0 \leq t \leq T_0$. The identified model can be used to predict how much of the share each opinion will have at a future time $T_1 (> T_0)$, and to identify both high decay tendency data sets and low decay tendency data sets.

3.1 Parameter Estimation

We describe a method for estimating decay parameter values of the TDV model from a given observed opinion spreading data \mathcal{D}_{T_0} . Based on the evolution process of our model (see Eq. (6)), we can obtain the likelihood function,

$$\mathcal{L}(\mathcal{D}_{T_0}; \lambda_J) = \log \left(\prod_{(k,t,v) \in \mathcal{D}_{T_0}} P(g_i(v) = k) \right), \quad (7)$$

where λ_J stands for the J -dimensional vector of decay parameter values, as explained in the previous subsection. Thus our estimation problem is formulated as a maximization problem of the objective function $\mathcal{L}(\mathcal{D}_{T_0}; \lambda_J)$ with respect to λ_J .

We derive an iterative algorithm for obtaining the maximum likelihood estimators. From the definitions of $P(g_i(v) = k)$ (see Eq. (6)) and $\rho(\Delta t; \lambda_J)$ (see Eq. (5)), we can express Eq. (7) as follows:

$$\begin{aligned} \mathcal{L}(\mathcal{D}_{T_0}; \lambda_J) = & \sum_{(k,t,v) \in \mathcal{D}_{T_0}} \log \left(1 + \sum_{u \in B(v)} \sum_{\tau \in M_k(t,u)} \exp(-\lambda_J^T F_J(t - \tau)) \right) \\ & - \sum_{(k,t,v) \in \mathcal{D}_{T_0}} \log \left(K + \sum_{u \in B(v)} \sum_{\tau \in M(t,u)} \exp(-\lambda_J^T F_J(t - \tau)) \right). \end{aligned} \quad (8)$$

Now, let $\bar{\lambda}_J$ be the current estimate of λ_J . We focus on the first term of the right-hand side of Eq. (8), and define $q_{k,t,v}(\tau; \lambda_J)$ by

$$q_{k,t,v}(\tau; \lambda_J) = \frac{\exp(-\lambda_J^T F_J(t - \tau))}{1 + \sum_{u \in B(v)} \sum_{\tau' \in M_k(t,u)} \exp(-\lambda_J^T F_J(t - \tau'))}$$

for any $k \in \{1, \dots, K\}$, $t \in (0, T]$, $v \in V$, and $\tau \in \bigcup_{u \in B(v)} M_k(t, u)$. Note that for any $(k, t, v) \in \mathcal{D}_{T_0}$,

$$q_{k,t,v}(\tau; \lambda_J) > 0, \quad \left(\forall \tau \in \bigcup_{u \in B(v)} M_k(t, u) \right), \quad (9)$$

$$\sum_{u \in B(v)} \sum_{\tau \in M_k(t, u)} q_{k,t,u}(\tau; \lambda_J) + \frac{1}{1 + \sum_{u \in B(v)} \sum_{\tau' \in M_k(t, u)} \exp(-\lambda_J^T \mathbf{F}_J(t - \tau'))} = 1. \quad (10)$$

We can transform our objective function as follows:

$$\mathcal{L}(\mathcal{D}_{T_0}; \lambda_J) = Q(\lambda_J; \bar{\lambda}_J) - \mathcal{H}(\lambda_J; \bar{\lambda}_J), \quad (11)$$

where $Q(\lambda_J; \bar{\lambda}_J)$ is defined by

$$\begin{aligned} Q(\lambda_J; \bar{\lambda}_J) = & - \sum_{(k,t,v) \in \mathcal{D}_{T_0}} \sum_{u \in B(v)} \sum_{\tau \in M_k(t, u)} q_{k,t,v}(\tau; \bar{\lambda}_J) \lambda_J^T \mathbf{F}_J(t - \tau) \\ & - \sum_{(k,t,v) \in \mathcal{D}_{T_0}} \log \left(K + \sum_{u \in B(v)} \sum_{\tau \in M(t, u)} \exp(-\lambda_J^T \mathbf{F}_J(t - \tau)) \right), \end{aligned} \quad (12)$$

and $\mathcal{H}(\lambda_J; \bar{\lambda}_J)$ is defined by

$$\begin{aligned} \mathcal{H}(\lambda_J; \bar{\lambda}_J) = & \sum_{(k,t,v) \in \mathcal{D}_{T_0}} \left\{ \sum_{u \in B(v)} \sum_{\tau \in M_k(t, u)} q_{k,t,v}(\tau; \bar{\lambda}_J) \log q_{k,t,v}(\tau; \lambda_J) \right. \\ & + \frac{1}{1 + \sum_{u \in B(v)} \sum_{\tau' \in M_k(t, u)} \exp(-\bar{\lambda}_J^T \mathbf{F}_J(t - \tau'))} \\ & \left. \times \log \left(\frac{1}{1 + \sum_{u \in B(v)} \sum_{\tau' \in M_k(t, u)} \exp(-\lambda_J^T \mathbf{F}_J(t - \tau'))} \right) \right\}. \end{aligned} \quad (13)$$

By Eqs. (9), (10), (13), and the property of the KL-divergence, it turns out that $\mathcal{H}(\lambda_J; \bar{\lambda}_J)$ is maximized at $\lambda_J = \bar{\lambda}_J$. Hence, we can increase the value of $\mathcal{L}(\mathcal{D}_{T_0}; \lambda_J)$ by maximizing $Q(\lambda_J; \bar{\lambda}_J)$ with respect to λ_J (see Eq. (11)).

We derive an update formula for maximizing $Q(\lambda_J; \bar{\lambda}_J)$. We focus on the second term of the right-hand side of Eq. (12) (see also the second term of the right-hand side of Eq. (8)), and define $r_{t,v}(\tau; \lambda_J)$ by

$$r_{t,v}(\tau; \lambda_J) = \frac{\exp(-\lambda_J^T \mathbf{F}_J(t - \tau))}{K + \sum_{u \in B(v)} \sum_{\tau' \in M(t, u)} \exp(-\lambda_J^T \mathbf{F}_J(t - \tau'))} \quad (14)$$

for any $t \in (0, T]$, $v \in V$, and $\tau \in \bigcup_{u \in B(v)} M(t, u)$. Note that for any $(k, t, v) \in \mathcal{D}_{T_0}$,

$$r_{t,v}(\tau; \lambda_J) > 0, \quad \left(\forall \tau \in \bigcup_{u \in B(v)} M(t, u) \right), \quad \sum_{u \in B(v)} \sum_{\tau \in M(t, u)} r_{t,u}(\tau; \lambda_J) < 1. \quad (15)$$

From Eqs. (12) and (14), we can easily see that the gradient vector of $Q(\lambda_J; \bar{\lambda}_J)$ with respect to λ_J is given by

$$\frac{\partial Q(\lambda_J; \bar{\lambda}_J)}{\partial \lambda_J} = - \sum_{(t,v,k) \in \mathcal{D}_{T_0}} \sum_{u \in B(v)} \left(\sum_{\tau \in M_k(t,u)} q_{t,v,k}(\tau; \bar{\lambda}_J) \mathbf{F}_J(t - \tau) - \sum_{\tau \in M(t,u)} r_{t,v}(\tau; \lambda_J) \mathbf{F}_J(t - \tau) \right). \quad (16)$$

Moreover, from Eqs. (14) and (16), we can obtain the Hessian matrix of $Q(\lambda_J; \bar{\lambda}_J)$ as follows:

$$\frac{\partial^2 Q(\lambda_J; \bar{\lambda}_J)}{\partial \lambda_J \partial \lambda_J^T} = - \sum_{(k,t,v) \in \mathcal{D}_{T_0}} \left\{ \sum_{u \in B(v)} \sum_{\tau \in M(t,u)} r_{t,v}(\tau; \lambda_J) \mathbf{F}_J(t - \tau) \mathbf{F}_J(t - \tau)^T - \left(\sum_{u \in B(v)} \sum_{\tau \in M(t,u)} r_{t,v}(\tau; \lambda_J) \mathbf{F}_J(t - \tau) \right) \left(\sum_{u \in B(v)} \sum_{\tau \in M(t,u)} r_{t,v}(\tau; \lambda_J) \mathbf{F}_J(t - \tau) \right)^T \right\}. \quad (17)$$

By Eq. (17), for any J -dimensional real column vector \mathbf{x}_J , we have

$$\begin{aligned} & \mathbf{x}_J^T \frac{\partial^2 Q(\lambda_J; \bar{\lambda}_J)}{\partial \lambda_J \partial \lambda_J^T} \mathbf{x}_J \\ &= - \sum_{(k,t,v) \in \mathcal{D}_{T_0}} \left\{ \sum_{u \in B(v)} \sum_{\tau \in M(t,u)} r_{t,v}(\tau; \lambda_J) (\mathbf{x}_J^T \mathbf{F}_J(t - \tau))^2 - \left(\sum_{u \in B(v)} \sum_{\tau \in M(t,u)} r_{t,v}(\tau; \lambda_J) \mathbf{x}_J^T \mathbf{F}_J(t - \tau) \right)^2 \right\} \\ &= - \sum_{(k,t,v) \in \mathcal{D}_{T_0}} \left\{ \sum_{u \in B(v)} \sum_{\tau \in M(t,u)} r_{t,v}(\tau; \lambda_J) \left(\mathbf{x}_J^T \mathbf{F}_J(t - \tau) - \sum_{u \in B(v)} \sum_{\tau' \in M(t,u)} r_{t,v}(\tau'; \lambda_J) \mathbf{x}_J^T \mathbf{F}_J(t - \tau') \right)^2 + \left(1 - \sum_{u \in B(v)} \sum_{\tau \in M(t,u)} r_{t,v}(\tau; \lambda_J) \right) \left(\sum_{u \in B(v)} \sum_{\tau \in M(t,u)} r_{t,v}(\tau; \lambda_J) \mathbf{x}_J^T \mathbf{F}_J(t - \tau) \right)^2 \right\}. \end{aligned}$$

Thus, by Eq. (15), we obtain

$$\mathbf{x}_J^T \frac{\partial^2 Q(\lambda_J; \bar{\lambda}_J)}{\partial \lambda_J \partial \lambda_J^T} \mathbf{x}_J \leq 0, \quad (\forall \mathbf{x}_J \in \mathbf{R}^J),$$

that is, the Hessian matrix is negative semi-definite. Hence, by solving the equation

$$\frac{\partial Q(\lambda_J; \bar{\lambda}_J)}{\partial \lambda_J} = \mathbf{0}_J$$

(see Eq. (16)), we can find the value of λ_J that maximizes $Q(\lambda_J; \bar{\lambda}_J)$. We employed a standard Newton Method in our experiments.

3.2 Model Selection

One of the important purposes of introducing the TDV model is to analyze how people are affected by their neighbors' past opinions for a specific opinion formation process. In what follows, for a given set of candidate decay functions (i.e., feature vectors), we consider selecting one being the most appropriate to the observed data \mathcal{D}_{T_0} of $|\mathcal{D}_{T_0}| = N$, where N represents the number of opinion manifestations by individuals.

As mentioned in Section 2, the base TDV model is a special TDV model equipped with the decay function that equally weights all the past opinions. Thus, we first examine whether or not the TDV model equipped with a candidate decay function can be more appropriate to the observed data \mathcal{D}_{T_0} than the base TDV model.³ To this end, we employ the likelihood ratio test. For a given feature vector $F_J(\Delta t)$, let $\hat{\lambda}_J(F_J)$ be the maximal likelihood estimator of the TDV model equipped with the decay function of $F_J(\Delta t)$. Since the base TDV model is the TDV model of $\lambda_J = \mathbf{0}_J$, the log-likelihood ratio statistic of the TDV model with $F_J(\Delta t)$ against the base TDV model is given by

$$Y_N(F_J) = \mathcal{L}(\mathcal{D}_{T_0}; \hat{\lambda}_J(F_J)) - \mathcal{L}(\mathcal{D}_{T_0}; \mathbf{0}_J). \tag{18}$$

It is well known that $2Y_N(F_J)$ asymptotically approaches to the χ^2 distribution with J degrees of freedom as N increases. We set a significance level α ($0 < \alpha < 1$), say $\alpha = 0.005$, and evaluate whether or not the TDV model with $F_J(\Delta t)$ fits significantly better than the base TDV model by comparing $2Y_N(F_J)$ to $\chi_{J,\alpha}$. Here, $\chi_{J,\alpha}$ denotes the upper α point of the χ^2 distribution of J degrees of freedom, that is, it is the positive number z such that

$$\frac{1}{\Gamma(J/2)2^{J/2}} \int_0^z y^{J/2-1} \exp\left(-\frac{y}{2}\right) dy = 1 - \alpha,$$

where $\Gamma(s)$ is the gamma function. We consider the set \mathcal{FV} of the candidate feature vectors (i.e., decay functions) selected by this likelihood ratio test at significance level α . Next, we find the feature vector $F_{J^*}(\Delta t) \in \mathcal{FV}$ such that it maximizes the log-likelihood ratio statistic $Y_N(F_J)$, ($F_J(\Delta t) \in \mathcal{FV}$), (see Eq. (18)), and propose selecting the TDV model equipped with the decay function of $F_{J^*}(\Delta t)$. If the set \mathcal{FV} is empty, we select the base TDV model for \mathcal{D}_{T_0} .

³ The base TDV model is not the only baseline model with which the proposed method is to be compared. The simplest one would be the random opinion model in which each user chooses its opinion randomly independent of its neighbors.

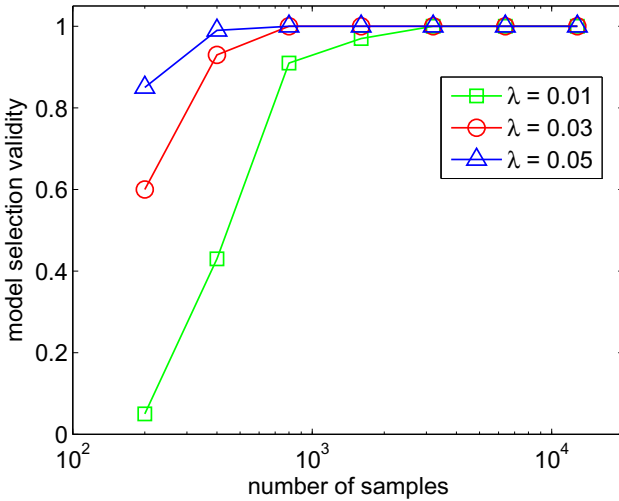


Fig. 1. Results of model selection validity for the exponential TDV model

Here we recall that typical decay functions in natural and social sciences include the exponential decay function (see Eq. (3)) and the power-law decay functions (see Eq. (4)). We refer to the TDV models of the exponential and the power-law decay functions as the *exponential TDV model* and the *power-law TDV model*, respectively. In our experiments, we in particular focus on investigating which of the base, the exponential, and the power-law TDV models best fits to the observed data \mathcal{D}_{T_0} . Thus, the TDV model to be considered has $J = 1$ and parameter λ .

4 Evaluation by Synthetic Data

Using synthetic data, we examined the effectiveness of the proposed method for parameter estimation and model selection. We assumed complete networks for simplicity. According to the TDV model, we artificially generated an opinion diffusion sequence \mathcal{D}_{T_0} consisting of 3-tuple (k, t, v) of opinion k , time t and node v such that $|\mathcal{D}_{T_0}| = N$, and applied the proposed method to the observed data \mathcal{D}_{T_0} , where the significance level $\alpha = 0.005$ was used for model selection. As mentioned in the previous section, we assumed two cases where the true decay follows the exponential distribution (see Eq. (3)) and the power-law distribution (see Eq. (3)), respectively. Let Y_N^e and Y_N^p denote the log-likelihood ratio statistics of the exponential and the power-law TDV models against the base TDV model, respectively (see Eq. (18)). We varied the value of parameter λ in the following range: $\lambda = 0.01, 0.03, 0.05$ for the exponential TDV model, and $\lambda = 0.4, 0.5, 0.6$ for the power-law TDV model, on the basis of the analysis performed for the real world @cosme dataset (see, Section 5). We conducted 100 trials varying the observed data \mathcal{D}_{T_0} of $|\mathcal{D}_{T_0}| = N$, and evaluated the proposed method.

First, we investigated the model selection validity $\mathcal{F}_N/100$, where \mathcal{F}_N is the number of trials in which the true model was successfully selected by the proposed method.

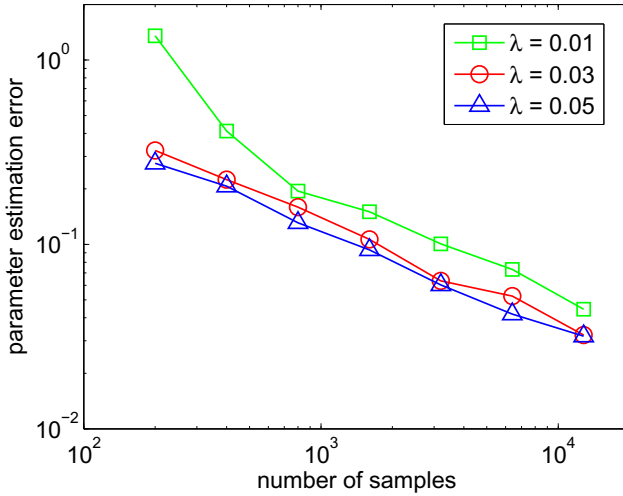


Fig. 2. Results of Parameter estimation error for the exponential TDV model

Namely, if the exponential TDV model is the true model, then \mathcal{F}_N is defined by the number of trials such that

$$2Y_N^e > \max(\chi_{1,\alpha}, 2Y_N^p),$$

and if the power-law TDV model is the true model, then \mathcal{F}_N is defined by the number of trials such that

$$2Y_N^p > \max(\chi_{1,\alpha}, 2Y_N^e).$$

Second, we examined the parameter estimation error \mathcal{E}_N for the trials in which the true model was selected by the proposed method. Here, \mathcal{E}_N is defined by

$$\mathcal{E}_N = \frac{|\hat{\lambda}(N) - \lambda^*|}{\lambda^*},$$

where λ^* is the true value of parameter λ , and $\hat{\lambda}(N)$ is the value estimated by the proposed method from the observed data \mathcal{D}_{T_0} of $|\mathcal{D}_{T_0}| = N$. Figures 1 and 2 show the results for the exponential TDV model, and Figures 3 and 4 show the results for the power-law TDV model. Here, Figures 1 and 3 display model selection validity $\mathcal{F}_N/100$ as a function of sample size N . Figures 2 and 4 display parameter estimation error \mathcal{E}_N as a function of sample size N . As expected, \mathcal{F}_N increases and \mathcal{E}_N decreases as N increases. Moreover, as λ becomes larger, \mathcal{F}_N increases and \mathcal{E}_N decreases. Note that a large λ means quickly forgetting past activities, and a small λ means slowly forgetting them. Thus, we can consider that a TDV model of smaller λ requires more samples to correctly learn the model. From Figures 1, 2, 3 and 4, we observe that the proposed method can work almost perfectly when N is greater than 500, and λ is greater than 0.01 for the exponential TDV model and greater than 0.4 for the power-law TDV model.

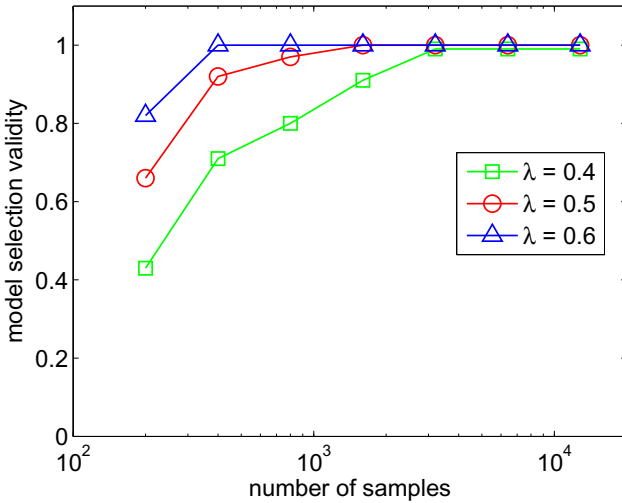


Fig. 3. Results of model selection validity for the power-law TDV model

5 Findings in Opinion Formation on Social Media

5.1 Dataset

We collected real data from “@cosme”⁴, which is a Japanese word-of-mouth communication website for cosmetics. In @cosme, a user can post a review and give a score of each brand (one from 1 to 7). When one user registers another user as his/her favorite user, a “fan-link” is created between them. We traced up to ten steps in the fan-links from a randomly chosen user in December 2009, and collected a set of (b, k, t, v) ’s, where (b, k, t, v) means that user v scored brand b k points at time t . The number of brands was 7,139, the number of users was 45,024, and the number of reviews posted was 331,084. For each brand b , we regarded the point k scored by a user v as the opinion k of v , and constructed the opinion diffusion sequence $\mathcal{D}_{T_0}(b)$ consisting of 3-tuple (k, t, v) . In particular, we focused on these brands in which the number of samples $N = |\mathcal{D}_{T_0}(b)|$ was greater than 500. Then, the number of brands was 120. We refer to this dataset as the @cosme dataset.

5.2 Results

We applied the proposed method to the @cosme dataset. Again, we adopted the temporal decay voter models with the exponential and the power-law distributions, and used the significance level $\alpha = 0.005$ for model selection. There were 9 brands such that $2Y_N^e > \chi_{1,\alpha}$, and 93 brands such that $2Y_N^p > \chi_{1,\alpha}$. Here, in the same way as the previous section, Y_N^e and Y_N^p denote the log-likelihood ratio statistics of the exponential and the power-law TDV models against the base TDV model, respectively.

⁴ <http://www.cosme.net/>

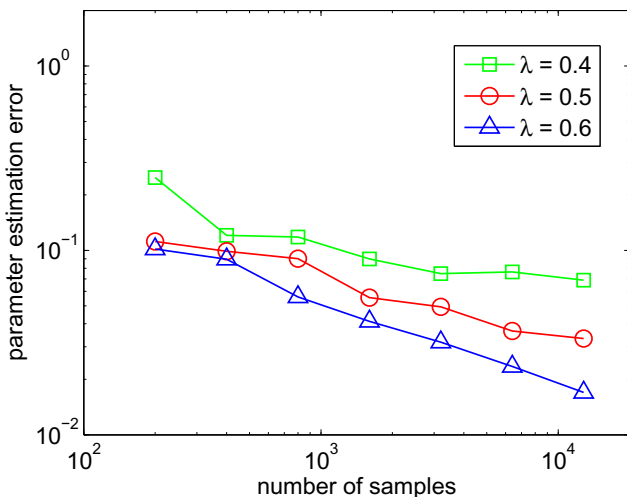


Fig. 4. Results of Parameter estimation error for the power-law TDV model

Further, there were 92 brands such that $2Y_N^p > \max(\chi_{1,\alpha}, 2Y_N^e)$, one brand such that $2Y_N^e > \max(\chi_{1,\alpha}, 2Y_N^p)$, and 27 brands such that $\max(2Y_N^p, 2Y_N^e) \leq \chi_{1,\alpha}$. Namely, according to the proposed method, 92 brands were the power-law TDV model, 27 brands were the base TDV model, and only one brand was the exponential TDV model. These results show that most brands conform to the power-law TDV model. This also agrees with the work [1,19] that many human actions are related to power-laws.

Figures 5 and 6 show the results for the @cosme dataset from the point of view of the power-law TDV model. Figure 5 plots the log-likelihood ratio statistic Y_N^p for each brand as a function of sample size N , where the thick solid line indicates the value of $\chi_{i,\alpha}$. In addition to the brands plotted, there is a brand such that $Y_N^p = Y_N^e = 0$. It was brand “YOJIYA”, which is a traditional Kyoto brand, and is known as a brand releasing new products less frequently. Thus, we speculate that it conforms to the base TDV model. Figure 6 plots the pair $(Y_N^p, \hat{\lambda}(N))$ for the brands in which the power-law TDV model was selected by the proposed method, where $\hat{\lambda}(N)$ is the value of parameter λ estimated by the proposed method from the observed data $\mathcal{D}_{T_0}(b)$ of $|\mathcal{D}_{T_0}(b)| = N$. From Figure 6, we observe that Y_N^p and $\hat{\lambda}(N)$ are positively correlated. This agrees with the fact that the power-law TDV model with $\lambda = 0$ corresponds to the base TDV model. In Figures 5 and 6, the big solid red circle indicates the brand “LUSH-JAPAN”, which had the largest values of Y_N^p , $\hat{\lambda}(N)$ and N , respectively. We also find the big solid green triangle in Figure 5 as a brand that had a large value of Y_N^p and a relatively small value of N . This was the brand “SHISEIDO ELIXIR SUPERIEUR”, which had the seventh largest value of Y_N^p , $N = 584$, and $\hat{\lambda}(N) = 0.58$. Note that these brands “LUSH-JAPAN” and “SHISEIDO ELIXIR SUPERIEUR” are known as brands that were recently established and release new products frequently. Thus, we speculate that they conform to the power-law TDV model with large λ .

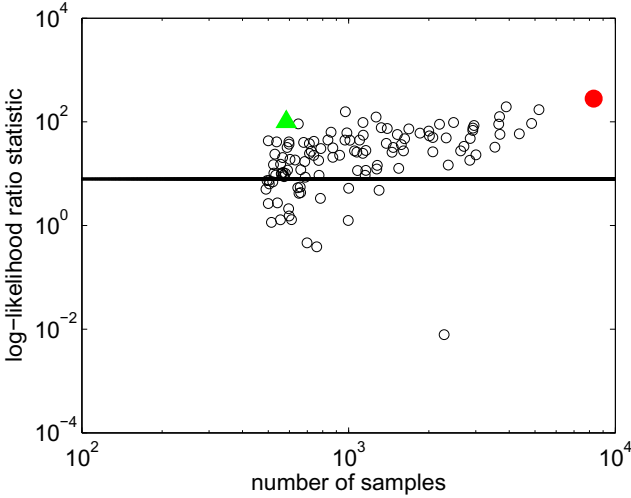


Fig. 5. Log-likelihood ratio statistic Y_N^p and number of samples N for the @cosme dataset

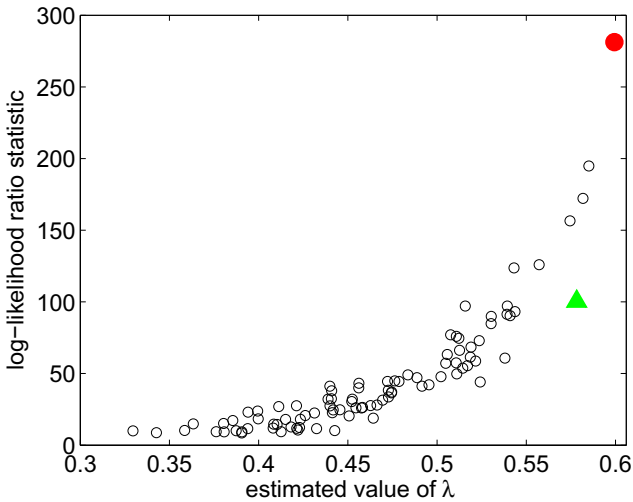


Fig. 6. Log-likelihood ratio statistic Y_N^p and estimated parameter value $\hat{\lambda}(N)$ for the @cosme dataset

6 Conclusion

We addressed the problem of how people make their own decisions based on their neighbors' opinions. The model best suited to discuss this problem is the voter model and several variants of this model have been proposed and used extensively. However, all of these models assume that people use their neighbors' latest opinions. People change opinions over time and some opinions are more persistent and some others

are less persistent. These depend on many factors but the existing models do not take this effect into consideration. In this paper, we, in particular, addressed the problem of how people's opinions are affected by their own and other peoples' opinion histories. It would be reasonable to assume that older opinions are less influential and recent ones are more influential. Based on this assumption, we devised a new voter model, called the temporal decay voter (TDV) model which uses all the past opinions in decision making in which decay is assumed to be a linear combination of representative decay functions each with different decay factors. The representative functions include the linear decay, the exponential decay, the power-law decay and many more. Each of them specifies only the form and the parameters remain unspecified. We formulated this as a machine learning problem and solved the following two problems: 1) Given the observed sequence of people's opinion manifestation and an assumed decay function, learn the parameter values of the function such that the corresponding TDV model best explains the observation, and 2) Given a set of decay functions each with the optimal parameter values, choose the best model and refute others. We solved the former problem by maximizing the likelihood and derived an efficient parameter updating algorithm, and the latter problem by choosing the decay model that maximizes the log likelihood ratio statistic. We first tested the proposed algorithms by synthetic datasets assuming that there are two decay models: the exponential decay and the power-law decay. We confirmed that the learning algorithm correctly identifies the parameter values and the model selection algorithm correctly identifies which model the data came from. We then applied the method to the real opinion diffusion data taken from a Japanese word-of-mouth communication site for cosmetics. We used the two decay functions above and added no decay function as a baseline. The result of the analysis revealed that opinions of most of the brands conform to the TDV model of the power-law decay function. We found this interesting because this is consistent with the observation that many human actions are related to the power-law. Some brands showed behaviors characteristic to the brands, e.g., the older brand that releases new product less frequently naturally follows no decay TDV and the newer brand that releases new product more frequently naturally follows the power-law decay TDV with large decay constant, which are all well interpretable.

Acknowledgments. This work was partly supported by Asian Office of Aerospace Research and Development, Air Force Office of Scientific Research under Grant No. AOARD-11-4111, and JSPS Grant-in-Aid for Scientific Research (C) (No. 23500194).

References

1. Barabási, A.L.: The origin of bursts and heavy tails in human dynamics. *Nature* 435, 207–211 (2005)
2. Castellano, C., Munoz, M.A., Pastor-Satorras, R.: Nonlinear q -voter model. *Physical Review E* 80, Article 041129 (2009)
3. Chen, W., Wang, Y., Yang, S.: Efficient influence maximization in social networks. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2009), pp. 199–208 (2009)

4. Chen, W., Yuan, Y., Zhang, L.: Scalable influence maximization in social networks under the linear threshold model. In: Proceedings of the 10th IEEE International Conference on Data Mining (ICDM 2010), pp. 88–97 (2010)
5. Crandall, D., Cosley, D., Huttenlocher, D., Kleinberg, J., Suri, S.: Feedback effects between similarity and social influence in online communities. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008), pp. 160–168 (2008)
6. Domingos, P.: Mining social networks for viral marketing. *IEEE Intelligent Systems* 20, 80–82 (2005)
7. Even-Dar, E., Shapira, A.: A Note on Maximizing the Spread of Influence in Social Networks. In: Deng, X., Graham, F.C. (eds.) WINE 2007. LNCS, vol. 4858, pp. 281–286. Springer, Heidelberg (2007)
8. Gruhl, D., Guha, R., Liben-Nowell, D., Tomkins, A.: Information diffusion through blogspace. *SIGKDD Explorations* 6, 43–52 (2004)
9. Holme, P., Newman, M.E.J.: Nonequilibrium phase transition in the coevolution of networks and opinions. *Physical Review E* 74, Article 056108 (2006)
10. Kempe, D., Kleinberg, J., Tardos, E.: Maximizing the spread of influence through a social network. In: Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2003), pp. 137–146 (2003)
11. Kimura, M., Saito, K., Nakano, R., Motoda, H.: Extracting influential nodes on a social network for information diffusion. *Data Mining and Knowledge Discovery* 20, 70–97 (2010)
12. Kimura, M., Saito, K., Ohara, K., Motoda, H.: Learning to predict opinion share in social networks. In: Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI 2010), pp. 1364–1370 (2010)
13. Kimura, M., Saito, K., Ohara, K., Motoda, H.: Detecting Anti-majority Opinionists Using Value-Weighted Mixture Voter Model. In: Elomaa, T., Hollmén, J., Mannila, H. (eds.) DS 2011. LNCS, vol. 6926, pp. 150–164. Springer, Heidelberg (2011)
14. Koren, Y.: Collaborative filtering with temporal dynamics. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2009), pp. 447–456 (2009)
15. Leskovec, J., Adamic, L.A., Huberman, B.A.: The dynamics of viral marketing. *ACM Transactions on the Web* 1, Article 5 (2007)
16. Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., Glance, N.: Cost-effective outbreak detection in networks. In: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2007), pp. 420–429 (2007)
17. Newman, M.E.J.: The structure and function of complex networks. *SIAM Review* 45, 167–256 (2003)
18. Newman, M.E.J., Forrest, S., Balthrop, J.: Email networks and the spread of computer viruses. *Physical Review E* 66, Article 035101 (2002)
19. Oliveira, J.G., Barabási, A.L.: Dawin and Einstein correspondence patterns. *Nature* 437, 1251 (2005)
20. Sood, V., Redner, S.: Voter model on heterogeneous graphs. *Physical Review Letters* 94, Article 178701 (2005)
21. Wu, F., Huberman, B.A.: How Public Opinion Forms. In: Papadimitriou, C., Zhang, S. (eds.) WINE 2008. LNCS, vol. 5385, pp. 334–341. Springer, Heidelberg (2008)
22. Yang, H., Wu, Z., Zhou, C., Zhou, T., Wang, B.: Effects of social diversity on the emergence of global consensus in opinion dynamics. *Physical Review E* 80, Article 046108 (2009)