The ‘hit’ phenomenon: a mathematical model of human dynamics interactions as a stochastic process

Akira Ishii\textsuperscript{1,3}, Hisashi Arakaki\textsuperscript{1}, Naoya Matsuda\textsuperscript{1}, Sanae Umemura\textsuperscript{1}, Tamiko Urushidani\textsuperscript{1}, Naoya Yamagata\textsuperscript{1} and Narihiko Yoshida\textsuperscript{1,2}

\textsuperscript{1} Department of Applied Mathematics and Physics, Tottori University, Koyama, Tottori 680-8552, Japan
\textsuperscript{2} Digital Hollywood University, Tokyo, Japan
E-mail: ishii@damp.tottori-u.ac.jp

\textit{New Journal of Physics} \textbf{14} (2012) 063018 (22pp)
Received 3 December 2011
Published 14 June 2012
Online at \url{http://www.njp.org/}
doi:10.1088/1367-2630/14/6/063018

\textbf{Abstract}. A mathematical model for the ‘hit’ phenomenon in entertainment within a society is presented as a stochastic process of human dynamics interactions. The model uses only the advertisement budget time distribution as an input, and word-of-mouth (WOM), represented by posts on social network systems, is used as data to make a comparison with the calculated results. The unit of time is days. The WOM distribution in time is found to be very close to the revenue distribution in time. Calculations for the Japanese motion picture market based on the mathematical model agree well with the actual revenue distribution in time.

\textsuperscript{3} Author to whom any correspondence should be addressed.
1. Introduction

Human interaction in ‘real’ society can be considered using the ‘many-body’ theory. With the popularization of social network systems (SNS) such as blogs, Twitter, Facebook, Google+ and other similar services around the world, interactions between accounts can be stored as digital data. Although the SNS society is not the same as real society, we can assume that communication is very similar. Thus, we can use the huge stock of human communication digital data as observation data for real society [1–4]. Using this observation, we can apply statistical mechanics methods to the social sciences. Since word-of-mouth (WOM) is very significant—for example, in marketing science [5–8]—such an analysis and prediction of the digital WOM in the sense of statistical physics is of importance today.

In this paper, as an applied field of the statistical mechanics of human dynamics, we focus our attention on motion picture entertainment, because the logs of communication for each movie in SNS can be distinguished easily, and the market competition between movies can be neglected because of the character of each movie; the markets for the Harry Potter series,
the *Pirates of the Caribbean* series and *Avatar* can be distinguished, for example. Moreover, traditionally, the motion picture industry in Japan has daily data on revenue for each movie.

The theoretical treatment of the motion picture business has a long history in the social sciences, such as marketing. The traditional method of forecasting motion picture revenue is to assume the following simple model:

\[ R = ABC e^D, \]  

where A, B, C and D represent qualities such as advertisement budget for the movie, strength of WOM, star power, quality of story, quality of music, etc. Then the formula is ‘linearized’ as follows:

\[ \log R = \alpha_1 \log A + \alpha_2 \log B + \alpha_3 \log D. \]  

Using the huge stock of market data, the coefficients \( \alpha_1, \alpha_2, \alpha_3 \) and \( \alpha_4 \) are determined using accurate statistics \([9–15]\). However, before discussing the actual determined coefficients, physicists must question the model of equation (1) itself. The form of equation (1) itself should be considered deeply and should be derived. Moreover, the model of equation (1) has no way to include the dynamics of human interactions because it is too simplified. In an actual society, including the SNS society, communication between humans has some dynamic behaviors, so that we should use a more realistic model to consider the aggregation behavior of communication in society.

Approaches from physicists also exist \([16–19]\). These approaches address the statistical law and the dynamics of motion picture popularity at the box office. Sinha and co-workers \([16]\) found the long-tail distribution of the popularity of top movies in theaters and discussed in detail the similarities and differences between two types of hits—blockbusters and sleepers—mainly in the US market. They suggested that popularity may be the outcome of a linear multiplicative stochastic process. They found the lognormal nature of the tail of total income and the bimodal form of the overall gross income distribution. They also discussed the nature of the decay of gross income per theater with time.

Before discussing their work, we should point out that the Japanese market is very different from the markets of the US and India. Because of the small land size and the high concentration of people in the metropolitan areas, the distribution of Japanese movie theaters is very concentrated in metropolitan areas such as the greater Tokyo area, the greater Osaka area and Nagoya, for example. Moreover, the three major motion picture companies, Toho, Toei and Shotiku, control most theaters in Japan. Because of this concentration of the location of theaters and the control of the major companies, sleeper-type hits never happen in the Japanese motion picture market. All Japanese hit movies are of the blockbuster type. Therefore, Sinha *et al*’s \([21]\) detailed analysis of sleeper-type hits cannot be applied to the Japanese motion picture entertainment industry.

For blockbusters, Pan and Sinha \([17]\) pointed out that the opening week is the most critical event in the commercial life of a movie. However, from their analysis of their weekly data, they concluded that advertising may not be a decisive factor in the success of a movie at the box office. Asur and Huberman \([18]\) used Twitter logs for movies, focusing their attention on the period comprising one week before the opening and the opening two weeks. However, they did not pay attention to the correlation between daily advertisements and daily weblog (blog) entries. Moreover, they did not pay attention to the daily advertisement budget, but only the total advertisement budget of a certain movie.
Ratkiewicz et al [19] discussed online popularity. They proposed a minimal model combining the classic preferential popularity increase mechanism with the occurrence of random popular shifts due to exogenous factors. They analyzed two large-scale networks: Wikipedia and the Chilean Web. Although their analysis is very interesting and useful, the sudden increase in the popularity of a certain movie has too short a duration for their approach.

The dynamics of the popularity increase have been investigated theoretically by presenting mathematical models to discuss it. The stochastic process has been tried as a way to forecast motion picture revenues [20], but the approach is still incomplete and could be made more accurate using data from blogs, Twitter or Facebook postings.

A better approach from the point of view of physicists is the so-called Bass model, which was presented as a simple model of aggregation behavior for WOM in 1969 [21, 22]. The key concept of the Bass model is a diffusion equation: diffusion of WOM in society. Many modified Bass models have been presented to analyze WOM for motion pictures [23, 24]. In the Bass model, we consider the number of adopters at the time \( t \), \( R(t) \). The number of non-adopters is calculated as \( N - R(t) \), where \( N \) is the number of persons in the market. If advertisements affect the number of people who adopt the products, we can write the increasing rate of adoption as

\[
\frac{dR(t)}{dt} = p(N - R(t)),
\]

(3)

where \( p \) is the probability that non-adopters will adopt the product per unit of time due to the advertisement. People can also be affected by WOM from the adopter. Thus, if we consider only the WOM effect, we find that

\[
\frac{dR(t)}{dt} = q(N - R(t))R(t),
\]

(4)

where \( q \) is the probability that non-adopters will adopt the product per unit time due to WOM from the adopters. Thus, combining both effects, we can write equation (5) as follows:

\[
\frac{dR(t)}{dt} = (N - R(t))(p + q R(t)).
\]

(5)

This is the equation of the Bass model.

In the Bass model, the advertisement is included only as the factor \( p \). The many modified Bass models include the decrease per time of the advertisement effect using the exponential decay function as follows:

\[
\frac{dR(t)}{dt} = (N - R(t)) \left( p e^{-\alpha(t-t_0)} + q R(t) \right),
\]

(6)

where \( t_0 \) is the time of the release day of the product. However, the real marketing actions begin several weeks before release. The modified Bass model above does not include such advertisement effects before release.

Moreover, the Bass and modified Bass models above do not include rumor effects in real society, that are not described using the person-to-person two-body interaction.

From the brief review of the previous studies above, we find that the effects of advertisements and WOM are included incompletely and the rumor effect is not included. Therefore, from the point of view of statistical physics, we present in this paper a model to include these three effects: the advertisement effect, the WOM effect and the rumor effect. The model presented is applied to the motion picture business in the Japanese market, and we compare our calculation with the reported revenue and observed number of blog postings for each film.
2. Model

2.1. Purchase intention

We start the modeling from the viewpoint of the individual consumer. We define the purchase intention of the individual consumer, labeled $i$, at time $t$ as $I_i(t)$. We assume that the number of products adopted until time $t$ can be written as

$$ Y(t) = p \int_{t_0}^{t} \sum_{i=1}^{N} I_i(t) \, dt, \quad (7) $$

where $N$ is the maximum number of adopted persons, $p$ is the price of the product and $t_0$ is the release date of the product. Thus, our problem is to define the equation of the purchase intention of each consumer $I_i(t)$. We consider the modeling of the effects of advertisements, WOM and rumor for the purchase intention in the following subsections.

2.2. Advertisement effect

The advertisement effect through mass media such as TV, newspapers, magazines, the Web, Facebook and Twitter is modeled as an external force for the equation of the purchase intention of the individual consumer:

$$ \frac{dI_i(t)}{dt} = c_i A(t), \quad (8) $$

where $A(t)$ is the time distribution of the effective advertisement effect per unit time and the coefficient describes the impression of the advertisement for consumer $i$. The external force $A(t)$ can be considered as trends in the world or political pressure on the market. In the application to the motion picture business in the Japanese market, we input the real daily advertisement budget used by the largest advertisement office in Japan, Dentsu Inc.

2.3. The word-of-mouth effect

Usually, a film’s success spreads through WOM. Such WOM sometimes has a very significant effect on the success of the movie. Thus, the WOM effect should be included in our theory.

The WOM effect can be distinguished into two types: WOM direct from friends and indirect WOM as rumors. We term the WOM effect between friends direct communication, because customers obtain information directly from their friends. In previous marketing theories based on the Bass model [21–24], usually only communications from adopter to non-adopter are taken into account. Here, we include communication between non-adopters. It is very significant for movie entertainment, especially before the opening of the movie. Let us consider that person $i$ hears information from person $j$. The probability per unit of time for the information to affect the purchase intention of person $i$ can be described as $D_{ij}I_j(t)$, where $I_j(t)$ is the purchase intention of person $j$ and $D_{ij}$ is the coefficient of the direct communication. A schematic image of the direct communication of persons $i$ and $j$ is shown in figure 1. Thus, we can write the effect of the direct communication as follows:

$$ \sum_{j=1}^{N} D_{ij}I_j(t), \quad (9) $$

where the summation is done without $j = i$. 

New Journal of Physics 14 (2012) 063018 (http://www.njp.org/)
In this paper, the rumor effect is termed as *indirect communication*. In this form of communication, a person hears a rumor while chatting on the street, overhearing a conversation from the next table in a restaurant or on a train, or finds the rumor in blogs or on Twitter. The situation is illustrated in figure 2(a), where many conversations are conducted on the streets of a city. To construct the theory using mathematics, we focus on one person who listens to a conversation happening around him/her. Let us consider that person $I$ overhears the conversation between person $j$ and person $k$. The strength of the effect of the conversation can be described as $D_{jk} I_j(t) I_k(t)$. The probability per unit time for the conversation to affect the purchase intention of person $i$ is defined as $Q_{ijk} D_{jk} I_j(t) I_k(t)$, where $Q_{ijk}$ is the coefficient. Thus, the indirect communication coefficient can be defined as $P_{ijk} = Q_{ijk} D_{jk}$. The situation is shown in figure 2(b).

Therefore, direct communication is two-body interaction and indirect communication is three-body interaction. Thus, our theory for ‘hit’ phenomena such as hit movies or music can
Figure 3. Movie revenues are shown as functions of date. Revenue data are the value in the Japanese market. The movie titles are as follows: (A) Death Note; (B) Death Note: The Last Name; (C) Zoku Always; (D) Close Zero; (E) Closed Note; (F) Biohazard 3; (G) Bushi no Ichibun; (H) Ratatouille; (I) fra-girl; (J) Dainihonjin; (K) Maiko haaaan!!; (L) Koizora; and (M) HERO.

be described as the equation of the purchase intention of person $i$ with two-body interaction and three-body interaction terms.

2.4. Decline of audience

A person tends to only watch a particular movie once. This means that the potential number of attendees decreases monotonically after the opening day. In figure 3, we show the typical decline of cinema box office returns for the Japanese movie market. It should be pointed out that DVD sales in Japan for a movie begin several months after the opening day at the cinema. This is the business practice in the Japanese entertainment industry. Thus, on the days shown in figure 3, no one can buy DVDs or online movies in Japan.

We found in figure 3 that the revenues decrease almost exponentially. This evidence is very natural, because the number of audience members decreases monotonically due to the effect that a person who has watched the movie does not watch the same movie again. In figure 3, only film E shows a sudden decrease. This happened due to a sudden scandal involving the actress in the movie. From the data, we find that the decay factors for most movies in figure 3 seem to be similar. The decay factor of the exponential decay is nearly 0.06 per day. This value is similar to the value reported in [17] for Spider-Man in the US market. It agrees well with the empirical rule of the Japanese movie market that the number of audience members undergoes roughly a 6% decrease.

This exponential decrease can be explained easily using a simple mathematical model. First, we denote the number of potential audience members as $N_0$ and the number of integrated audience members at time $t$ as $N(t)$. Thus, the number of people who are interested in the movie...
is \( N_0 - N(t) \). Assuming that the probability of watching the movie per day is \( a \), we obtain the equation to describe the number of audience members as

\[
\frac{dN(t)}{dt} = a(N_0 - N(t)).
\]

The solution of the equation with \( N_0 = 0 \) at \( t = 0 \) is

\[
N(t) = N_0(1 - e^{-at}).
\]

It is clear that the result can explain roughly the exponential decay of the audience shown in figure 3. The purchase intention of person \( i \), \( I_i(t) \), can also be considered to decay in a similar manner.

2.5. Equation of purchase intention for the ‘hit’ phenomenon

According to the above consideration, we write down the equation of purchase intention at the individual level as

\[
\frac{dI_i(t)}{dt} = -aI_i(t) + \sum_j d_{ij}I_j(t) + \sum_j \sum_k h_{ijk}I_j(t)I_k(t) + f_i(t),
\]

where \( d_{ij}, h_{ijk} \) and \( f_i(t) \) are the coefficient of the direct communication, the coefficient of the indirect communication and the random effect for person \( i \), respectively. We consider the above equation for every consumer so that \( i = 1, \ldots, N_p \).

Taking into account the effect of direct communication, indirect communication and the decline of audience, we obtain the above equation for the mathematical model for the hit phenomenon. The advertisement and publicity effect for each person can be described as the random effect \( f_i(t) \).

Equation (12) is the equation for all individual persons, but it is not convenient for analysis. Thus, we consider the ensemble average of the purchase intention of individual persons as follows:

\[
I(t) = \frac{1}{N} \sum_i I_i(t).
\]

Taking the ensemble average of equation (12), we obtain for the left-hand side

\[
\frac{dI(t)}{dt} = \frac{1}{N} \sum_i \frac{dI_i(t)}{dt} = \frac{d}{dt} \left( \frac{1}{N} \sum_i I_i(t) \right) = \frac{dI}{dt}.
\]

For the right-hand side, the ensemble average of the first, second and third terms is as follows:

\[
-aI_i = -a \frac{1}{N} \sum_i I_i(t) = -aI(t),
\]

\[
\sum_j d_{ij}I_j(t) = \sum_j dI_j(t) = \frac{1}{N} \sum_i \sum_j dI_j(t) = \sum_i \frac{1}{N} \sum_j I_j(t) = NdI(t),
\]
\[
\sum_j \sum_k p_{ijk} I_j(t) I_k(t) = p \sum_j \sum_k I_j(t) I_k(t)
\]
\[
= \frac{1}{N} \sum_i p \sum_j \sum_k I_j(t) I_k(t)
\]
\[
= \sum_i p \frac{1}{N} \sum_j \sum_k I_j(t) I_k(t)
\]
\[
= Np \sum_i \frac{1}{N} \sum_j I_j(t) \frac{1}{N} \sum_k I_k(t)
\]
\[
= N^2 p I(t)^2,
\] (17)
where we assume that the coefficient of direct and indirect communications can be approximated to be
\[
d_{ij} \cong d,
\]
\[
h_{ijk}d_{jk} = p_{ijk} \cong p
\]
under the ensemble average.

For the fourth term, the random effect term, we consider that the random effect can be divided into two parts: the collective effect and the individual effect:
\[
f_i(t) = f(t) + \Delta f_i(t),
\] (18)
\[
f_i(t) = \frac{1}{N} \sum_i f_i(t) = f(t),
\] (19)
where \(\Delta f_i(t)\) means the deviation of the individual external effects from the collective effect, \(f(t)\). Thus, we consider here that the collective external effect term \(f(t)\) corresponds to advertisements and publicity to persons in society. The deviation term \(\Delta f_i(t)\) corresponds to the deviation effect from the collective advertisement and publicity effect for individual persons, which we can assume to be
\[
\Delta f_i(t) = \frac{1}{N} \sum_i \Delta f_i(t) = 0.
\] (20)

Therefore, we obtain the equation for the ensemble-averaged purchase intention in the following manner:
\[
\frac{dI(t)}{dt} = -aI(t) + DI(t) + PI(t)^2 + f(t),
\] (21)
where
\[
Nd = D,
\]
\[
N^2 p = P.
\] (22)

Equation (21) can be applied to the purchase intention in the real market. Equation (21) is the equation we assumed in our previous works without derivation [25–30]. In this paper, we apply this equation to the motion picture business.
3. Observed data in the market

For the application of equation (21) to the real markets of motion pictures, we observe some market data as inputs and compare the observation with our calculation. The market data we use here are daily advertisement cost, daily revenue and daily number of blog postings on the Internet.

3.1. Daily advertisement data

In much of the previous marketing science research [1–15, 20–24], advertising costs are considered only in total. However, the time distribution of the daily advertising cost is very important. In figure 4, we show the advertising cost and related gross rating point (GRP) for *The Da Vinci Code* in the Japanese market. GRP is a term used in advertising to measure the size of the audience reached by a specific media vehicle or schedule [31]. It is just the product of the percentage of the target audience reached by an advertisement times the frequency at which they see it in a given movie campaign. Both the daily advertising costs and the GRP for each movie in the Japanese market are obtained from Dentsu Inc., the largest advertising agency in Japan. In Japan, Dentsu handles most advertisements appearing on every TV station except NHK (the national station), so all campaigns for all movies in all Japanese prefectures are organized entirely by Dentsu; thus, the advertisement cost data from Dentsu are exactly equal to the advertisement costs of the movies in the Japanese market. The value of the daily advertising costs does not include the discount for major movie production agencies in Japan, but the discount rate is the same during the campaign of each movie.

It was clearly seen in figure 4 that the advertising costs and GRP are distributed for two months after the opening of a movie. This feature can be found for every movie campaign in the Japanese market. Thus, not only the total advertising cost but also the time distribution of the costs is very important for the success of the advertising campaign for a motion picture.
3.2. Daily audience

The number of daily audience members for each movie is very important for the mathematical study of a motion picture campaign. In Japan, sales of DVDs or online access to movies is forbidden until several months after the opening of the movie in movie theaters. Thus, the number of audience members is the exact number of people who watch the movie. During this time, no one can watch the movie without going to the movie theater. We obtain the data for daily audience numbers from Box Office Japan (Kogyo Tsushinsha, Tokyo).

3.3. Daily blog postings

Daily blog postings for a movie are a very important signal to measure the movement of purchase intention among persons in the society. We measure the daily data of the number of posts for movies using the site Kizasi, which is a service for observing blog postings in Japan. We measure the number of blog posts for 25 movies in Japan in order to compare this information with the box office gross income for each movie in Japan. We show several results in figures 5–9. In these figures, the curve of daily blog posts is multiplied by a certain constant to normalize the value to be similar to the revenue value.

From figures 5–9, we find that the curve of the daily blog posts for each movie is very similar to the curve of the daily revenue we reported in our previous works [27, 28]. We observed 25 movies in the Japanese market and found a similar feature for all 25 movies. Since the selected 25 movies were a random sampling of hit movies in Japan, we can consider that the similarity can be applied to the most popular motion pictures in the Japanese market. This observation means that the ratio of blog postings for each person is almost constant during the duration of each movie so that the daily number of blog postings is very similar to the daily revenue.

Blog postings for each film can be distinguished into positive, negative and neutral opinions. A positive opinion means that the blogger wants to watch the film or judges the watched film in a positive way. In figure 10, we show that more than half of the blogs show...
a positive opinion for several movies. Moreover, we find that the ratio of positive, negative and neutral opinions is almost constant during the duration of the movie opening. Thus, the observed blog posting counts can be considered to be proportional to the counts of positive blog posts.

According to this observation, we propose to use the daily number of blog posts as the daily ‘quasi-revenue.’ Quasi-revenue is very useful for analysis, because it can be defined even before the opening of the movie. We can observe the increase in anticipation of a movie.

4. Calculation

We calculate the daily purchase-intention using equation (21) with the daily advertisement cost as input data for $\langle f(t) \rangle$. The parameters in equation (21) are adjusted to fit the data with the observed quasi-revenue for each movie.
4.1. Calculation in detail

For the calculation of each movie, we should derive the actual formulation for the calculation where the adopter and the non-adopter are distinguished. We can write the direct communication term of the non-adopter-to-non-adopter interaction in equation (21) as follows:

\[
DI(t) = \frac{1}{N} \sum_i \sum_j dI_j(t)
= \frac{1}{N} d \sum_i N \frac{1}{N} \sum_j I_j(t)
= \frac{N_p - N(t)}{N_p} (N_p - N(t)) \frac{1}{N_p - N(t)} \sum_j d_{nn} I_j(t)
= \frac{N_p - N(t)}{N_p} (N_p - N(t)) d_{nn} I(t),
\]

(23)
Figure 10. Positive opinion ratio for Japanese movie blogs as a function of time measured from the opening day for each movie.

where

$$N(t) = N_p \int_0^t I(\tau) \, d\tau.$$  \hfill (24)

The direct communication term for the adopter to the non-adopter is

$$DI(t) \Rightarrow \frac{N(t)}{N_p} (N_p - N(t)) \, d_{\text{ad}} I.$$  \hfill (25)

Similarly, we obtain the indirect communication term due to the communication between non-adopters at time $t$:

$$PI(t)^2 = p \frac{1}{N} \sum_i \sum_j \sum_k I_j(t)I_k(t)$$

$$= p \frac{1}{N} \sum_i N \frac{1}{N} \sum_j N \frac{1}{N} \sum_k I_j(t)I_k(t)$$

$$= \left( \frac{N_p - N(t)}{N_p} \right)^3 N_p^2 \, p_{\text{ad}} I^2;$$

$$\frac{(N_p - N(t))^3}{N_p} \, p_{\text{ad}} I^2,$$  \hfill (26)

where $p_{\text{ad}}$ is the factor of the indirect communication between non-adopters at time $t$. For the indirect communication, we obtain two more terms corresponding to the indirect communication due to the communication between adopters and that between an adopter and a non-adopter as follows:

$$\frac{(N(t))^2 (N_p - N(t))}{N_p} p_{\text{yy}} I^2 + \frac{N(t) (N_p - N(t))^2}{N_p} p_{\text{ny}} I^2,$$  \hfill (27)

where $p_{\text{yy}}$ is the factor of the indirect communication between adopters and $p_{\text{ny}}$ is the factor of the indirect communication due to the communication between adopters and non-adopters at time $t$.  

New Journal of Physics 14 (2012) 063018 (http://www.njp.org/)
Finally, we obtain the equation of purchase intention for the actual calculation as follows:

\[
\frac{dI(t)}{dt} = -aI(t) + f(t) + \frac{N_p - N(t)}{N_p} (N_p - N(t)) d^{nn} I + \frac{N(t)}{N_p} (N_p - N(t)) d^{ny} I \\
+ \frac{(N_p - N(t))^2}{N_p} p^{nn} I^2 + \frac{(N(t))^2 (N_p - N(t))}{N_p} p^{ny} I^2 \\
+ \frac{N(t) (N_p - N(t))^2}{N_p} p^{nn} I^2.
\] (28)

For the period before the opening of the movie, there are no adopters. Furthermore, the total number of adopters is zero. Thus, for the period before opening, equation (28) is reduced to the following form:

\[
\frac{dI(t)}{dt} = -aI(t) + f(t) + N_p d^{nn} I(t) + N_p^2 p^{nn} I(t)^2.
\] (29)

Equation (28) with (24) is the nonlinear integro-differential equation. However, since the data are handled daily, the time difference is 1 day. We can solve the equation numerically as a difference equation.

4.2. Reliable factor

For the purpose of reliability, we introduce here the so-called ‘R-factor’ (reliable factor), well known in the field of low-energy electron diffraction (LEED) experiments [32]. In LEED experiments, the experimentally observed curve of current versus voltage is compared with the corresponding theoretical curve using the R-factor.

For our purpose, we define the R-factor as follows:

\[
R = \frac{\sum_i (f(i) - g(i))^2}{\sum_i (f(i)^2 + g(i)^2)},
\] (30)

where the functions f(i) and g(i) are defined in figure 11. The smaller the R-factor, the better the match in functions f and g. We use this R-factor as a guide to obtain the best adjustment of our parameters for each movie.

4.3. Results

We perform calculations using equation (29) for many movies in the Japanese market. We show several results in figures 12–16. The solid line shows our calculation with the real daily advertisement costs as input data of \( \langle f(t) \rangle \) of equation (29). In the calculation, we use trial and error to decide the parameters to minimize the R-factor. The durations we calculate the R-factors are equal to those of each corresponding figure. We also perform Metropolis-like random number sampling for some parameters to minimize the R-factor, but not for all the parameters, because it would require too much calculation time. The R-factor for each calculation is shown
in the figure captions of figures 12–16. For $D_{an}$ and $P_{an}$, the coefficients for direct and indirect communication between non-adopters, we decide the values before and after the opening day separately, because the audience response can change from the impression of the movie itself.

The results are compared with the number of blog posts as the quasi-revenue shown as a red histogram in the figures. We found that the agreement of the calculation with the quasi-revenue (blog posts) is very good.

The parameters we use in the above calculations are shown in table 1. We found that the value is similar for these movies, although not equal. The value for Avatar will be discussed later.
5. Discussion

5.1. Agreement with observations

In figures 12–16, we find that our calculations agree well with the observed daily number of blog posts for each movie. Since we find that the daily number of blog posts for a movie is proportional to the revenue for the movie, this agreement means that our calculation can reproduce the revenue for each movie. For the Japanese motion picture market, new movies are...
watched in cinemas, because online downloads and DVDs are available only several months after the opening day of the movie. Moreover, the cinema entrance fees in the Japanese market are similar for almost all cinemas. Thus, the agreement of our calculation with the daily blog post numbers means that our calculation can describe the movement of people in society very well, at least for the Japanese motion picture market.

The agreement here can be considered to mean that our theory can describe the collective motion of people in society, at least for the entertainment market in Japan. Since our theory is considered to be general, we expect that it can be applied to other entertainment markets in the world.
Table 1. Parameters used in the calculation.

|                  | Pirates of the Caribbean 3 | The Da Vinci Code | Close Zero | Always | Avatar (best) |
|------------------|-----------------------------|-------------------|------------|--------|---------------|
| \( C_{\text{adv}} \) before opening | 0.0129                     | 0.016             | 0          | 0.01   | 0.03          |
| \( C_{\text{adv}} \) after opening  | 0.0179                     | 0.02              | 0          | 0.0179 | 0             |
| \( a \)          | 0.3                        | 0.3               | 0.3        | 0.3    | 0.1           |
| \( D_{\text{nn}} \) before opening | 0.0179                     | 0.0179            | 0.0571     | 0.0179 | 0.13          |
| \( D_{\text{nn}} \) after opening  | 0.25                       | 0.299             | 0.175      | 0.04   | 0.3           |
| \( D_{\text{ny}} \)                  | 0.0599                     | 0.08              | 0          | 0.14   | 0.2           |
| \( P_{\text{nn}} \) before opening | 9.99E−07                   | 9.99E−07          | 9.99E−06   | 9.99E−07 | 1.70E−06     |
| \( P_{\text{nn}} \) after opening   | 1.80E−05                   | 2.49E−05          | 1.10E−07   | 8.50E−06 | 1.00E−05     |
| \( P_{\text{ny}} \)                  | 4.99E−07                   | 4.99E−07          | 0.00E+00   | 2.20E−07 | 1.00E−04     |
| \( P_{23} \)                     | 9.99E−06                   | 1.49E−05          | 0.00E+00   | 1.59E−05 | 6.50E−06     |

5.2. Optimization of time distribution of advertisement cost

As we saw in section 3.1, the advertisement cost has a distribution in time. The calculated result using equations (28) and (29) of our theory depends strongly on the time distribution of the advertisement cost. This means that, even for the same total cost, the total revenue can change depending on the time distribution of the advertisement cost. Thus, using our equations (28) and (29), we can consider how to optimize the time distribution of the advertisement cost to obtain the maximum value of the total revenue at the condition of the fixed total advertisement cost.

5.3. Indirect communication

One of the original ideas of the present theory is the effect of indirect communication, as explained in section 2.3. To demonstrate the effect of indirect communication, we perform two calculations: one including indirect communication and one without indirect calculation. The example we show here is the calculation for Avatar, which was a very big hit in 2010. Both calculations are optimized as much as possible using the Metropolis-like method of parameter optimization with random numbers to minimize the \( R \)-factor. The result is shown in figure 17. We found that the indirect calculation is better. The parameters for the best fit are shown in table 1. As we can see in table 1, the parameters of indirect communication are larger for Avatar. The result shows us that indirect communication played a very significant role in causing the success of Avatar. Since this calculation can be applied to many other big hit movies, we can conclude that indirect communication as included in our theory is very important in describing human interactions in real society.

5.4. Relation to the Bass model

The Bass model \([21–22]\) is well known to be the model of the spread of WOM. In this subsection, we verify that the Bass model is derived from the equation of the mathematical model of the hit phenomenon. For sales of the production, the equation of our mathematical
Figure 17. The indirect communication effect for *Avatar*. Day 0 is the opening day of the movie *Avatar* in Japan. ‘adv’ means the daily advertisement cost, ‘blog’ means the daily number of blog postings, ‘direct’ means the calculation excluding the indirect communication effect and ‘indirect’ means the calculation including indirect communication. For both calculations, the parameters in the calculation are optimized using the $R$-factor.

model for the hit phenomenon is written as follows:

$$\frac{dI_i(t)}{dt} = A_i + \sum_j D_{ij} I_j(t) + \sum_j \sum_k P_{ijk} I_j(t) I_k(t).$$  \hspace{1cm} (31)

The total number of sales for the product, $N$, can be defined using the number of people, $m$, and the purchase intention, as follows:

$$\sum_i I_i(t) = m I(t) = N(t).$$  \hspace{1cm} (32)

In the Bass model, advertisement is spread from adopters to non-adopters. Thus,

$$\sum_i A_i = a (m - N(t)) = \sum_i a,$$  \hspace{1cm} (33)

where the sum is only the persons who do not buy the product.

For the direct communication term in equation (31), $\sum_i \sum_j D_{ij} I_i I_j$, we assume here that the coefficient $D_{ij}$ is not zero only for the pair of persons, adopter and non-adopter. The purchase intention for the adopter is considered to be $I_j = 0$. Thus, we can express the direct communication term as follows:

$$\sum_i \sum_j D_{ij} I_j = (m - N(t)) N(t) \frac{dI}{dt} = (m - N(t)) N(t) b.$$  \hspace{1cm} (34)
Thus, summing up for $i$, we obtain from equation (21)

$$\frac{d}{dt} \sum_i I_i(t) = \sum_i \left[ A_i + \sum_j D_{ij} I_j(t) + \sum_j \sum_k P_{ijk} I_j(t) I_k(t) \right].$$

(35)

Therefore, substituting (34) into (28), we obtain

$$\frac{dN(t)}{dt} = a(m - N(t)) + (m - N(t)) N(t)b + \sum_i \sum_j \sum_k P_{ijk} I_j(t) I_k(t).$$

(36)

Here, if we neglect the indirect communication term as $P_{ijk} = 0$, we obtain the well-known Bass model equation

$$\frac{dN(t)}{dt} = a(m - N(t)) + (m - N(t)) N(t)b.$$

(37)

Therefore, we find that our mathematical model for the hit phenomenon includes the Bass model and addresses the previously neglected indirect communication. Thus, the indirect communication in equation (21) is the new WOM effect.

5.5. Statistical physics method for human interactions

In this paper, we calculate the time variation of the averaged action of humans in real society using our theory. Equation (21) or equations (28) and (29) seem to describe the collective action of humans for entertainment well. As in the usual physics approach, we calculate the prediction and compare it with observation. Thus, if we obtain a large quantity of observation data from real society, the usual method of statistical physics or many-body physics is available to investigate the human interaction in real society. For this purpose, of course, the human interaction for each individual problem should be investigated carefully with a social scientist specialising in communication.

6. Conclusion

We present the mathematical model of the hit phenomenon as an equation of consumer action where consumer–consumer communication is taken into account. In the communication effect, we include both direct communication and indirect communication. We found the daily number of blog posts to be very similar to the revenue of the corresponding movie. The daily number of blog posts can be used as quasi-revenue. The results calculated with the model can predict the revenue of the corresponding movie very well. We found that indirect communication affects revenue using the calculation and our theory. The conclusion presented in this paper will be applicable to any consumer market.

Acknowledgments

Dentsu Inc. provided the advertisement cost data. The authors are grateful to Takeshi Nakagawa of Dentsu Inc. for helping us to use the data. The revenue data were provided by Nikkan Kogyo Tsushinsya. This research was partially supported by the Hit Contents Laboratory Inc. The Okawa Foundation supports research on information and telecommunications. The authors are also grateful for the illustrations in figures 1 and 2 by Ms Nobue Inomoto, the secretary of the author Akira Ishii.
References

[1] Allsop D T, Basset B R and Hoskins J A 2007 J. Adv. Res. 47 398
[2] Kostka J, Oswald Y A and Wattenhofer R 2008 Lect. Notes Comput. Sci. 5058/2008 185–96
[3] Bakshy E, Hofman J M, Mason W A and Watts D J 2011 Proc. Fourth ACM Int. Conf. on Web Search and Data Mining doi:10.1145/1935826.1935845
[4] Jansen B J, Zhang M, Sobel K and Chowdury A 2009 J. Am. Soc. Inf. Sci. Technol. 60 2169
[5] Brown J J and Reingen P H 1987 J. Consum. Res. 14 350
[6] Murray K 1991 J. Mark. 55 10
[7] Banerjee A 1992 Q. J. Econ. 107 797
[8] Taylor J 2003 Brandweek 44 26
[9] Elberse A and Eliashberg J 2003 J. Mark. Sci. 22 329
[10] Liu Y 2006 J. Mark. 70 74
[11] Duan W, Gu B and Whinston A B 2008 Decis. Support Syst. 45 1007
[12] Duan W, Gu B and Whinston A B 2008 J. Retailing 84 233
[13] Zhu M and Lai S 2009 Proc. Int. Conf. on Electronic Commerce and Business Intelligence 120–4
[14] Goel S, Hofman J M, Laaha S, Pennock D M and Watts D J 2010 Proc. Natl Acad. Sci. USA 107 1786
[15] Karniouchina E V 2011 Int. J. Res. Mark. 28 62
[16] Sinha S and Raghavendra S 2004 Eur. Phys. J. B 42 293
[17] Pan R K and Sinha S 2010 New J. Phys. 12 115004
[18] Asur S and Huberman R A 2010 arXiv:1003.5699v1
[19] Ratkiewicz J, Fortunato S, Flammini A, Menczer F and Vespignani A 2010 Phys. Rev. Lett. 105 158701
[20] Eliashberg J, Jonker J-J, Sawhney M S and Wierenga B 2000 Mark. Sci. 19 226
[21] Bass F M 1969 Manage. Sci. 15 215–27
[22] Bass F M 1986 The adoption of a marketing model: comments and observations Innovation Diffusion Models of New Product Acceptance ed V Mahajan and Y Wind (Cambridge, MA: Ballinger)
[23] Dellarocas C, Awad N F and Zhang X 2004 Working Paper MIT Sloan School of Management
[24] Dellarocas C, Zhang X and Awad N F 2007 J. Interact. Mark. 21 23–45
[25] Ishii A and Yoshida N 2005 Reports of the Faculty of Engineering vol 36, Tottori University p 71 (in Japanese)
[26] Ishii A, Yoshida N, Arakaki H and Yamazaki F 2007 Reports of the Faculty of Engineering vol 37, Tottori University p 107 (in Japanese)
[27] Ishii A, Umemura S, Hayashi T, Matsuda N, Nakagawa T, Arakaki H and Yoshida N 2010 arXiv:1002.4460
[28] Yoshida N, Ishii A and Arakaki H 2010 Daihitto no Houteishiki (Equation of Big Hit) Discover Twenty One, Tokyo, 201, ISBN: 4887598467 (in Japanese)
[29] Ishii A, Arakaki H, Umemura S, Urushidani T, Matsuda N and Yoshida N 2011 Proc. 2011 Int. Conf. on Data Engineering and Internet Technology (DEIT 2011) (Conference Record No. 17831, IEEE Catalog No. CFP1113L-CDR, ISBN 978-1-4244-8581-9)
[30] Ishii A, Matsumoto T and Miki S Prog. Theor. Phys. at press
[31] Berkovitz T 2011 Political Media Buying: A Brief Guide Campaign '96: Third-Party Time harvard.edu (retrieved 16 June 2011)
[32] Pendry J B 1980 J. Phys. C: Solid State Phys. 3 937

New Journal of Physics 14 (2012) 063018 (http://www.njp.org/)