Prediction of Political Elections from Social Media Postings
Mathematical Model of Hit Phenomena

Akira Ishii[1], Yosuke Hayashi[1], Hidehiko Koguchi[2,3], Shitaro Takahashi[4] and Koki Uchiyama[5]
1 Department of Applied Mathematics and PhysicsTottori University, Koyama, Tottori 680-8554, Japan
2 Perspective Media, 3-6-9 Kakinokizaka, Meguro-ku, Tokyo 152-0022 Japan
3 M Data Co.Ltd., Toranomon, Minato-ku, Tokyo 105-0001 Japan
4 Digital Hollywood University, Sotokanda, Chiyoda-ku, Tokyo 101-0021, Japan
5 Hottolink, Kanda-nishikicho, Chiyoda-ku, Tokyo 101-0054, Japan

***Abstract*—Using the mathematical model for hit phenomena, we analyze the political elections in Japan. The mathematical model of hit phenomena is a theory of time-dependent human-human interaction based on the many-body theory in physics. In contrast to the simple measurement of the total number of posting to weblogs and twitter for each candidate of elections, we found that the direct communication strength and the indirect communication strength defined in the mathematical model for hit phenomena play significant role in the analysis of the daily number of posting to weblogs and twitter. The order of voting for each political elections in Japan can be reproduced exactly using our model for the Oaka mayor election in 2011 and the election of the Osaka area of the House of Councillors of Japan in 2010.**

Introduction

Today, social media becomes very popular so that every social activities are affected at least partially from social media like weblogs, Twitter, Facebook and Google+. For economic activities, social media should be considered in marketing. For political activities, the effect of social media is also considered to be very strong not only the Arab Spring in 2010-2011 but many other countries in the world[1,2]. Especially for political elections, the effect of social media is very significant to predict the result of the elections before the pol. For the Japanese election we concern, there are some studies for the correlation between political election and social meda[3].

In this paper, we apply the mathematical model for hit phenomena to analyze the effect of social media to the political elections in Japan. Our approach can be used for the prediction of results of political elections.

The mathematical model of hit phenomena we used here is presented by one of the author for the prediction of hit phenomena in entertainment businesses[4-9]. Hit phenomena in entertainments are very popular dynamical phenomena in real economics. For analysis of hit phenomena, we consider it as time-dependent non-equilibrium phenomena. In the mathematical model of hit phenomena, we calculate the purchase intention of each consumers as the result of the effect of advertisements, direct communication among consumers and indirect communication or rumor as an solution of differential equation for time. For the purpose of the present paper, we apply the similar equation for vote intention.

In this study, we consider the hit phenomena of political votes with both experimental and theoretical ways. For experimental viewpoint, we observe daily weblog-posting data and twitter-posting data for several candidates in each Japanese political election using the Hottolink Co.Ltd, a social media monitoring platform vendor in Japan. We also obtain the daily exposure for each candidate on television using the database presented by M Data Co. Ltd. in Japan. For theoretical viewpoint, we present here a mathematical model for hit phenomena as microscopic voter model where we treat the TV exposure as an external force from the candidate side and the voter to voter interaction as many body interaction just like many-body theory in solid state physics. In the present model, equation of voter action is presented for each voter persons.

Observation of election

We observe three Japanese political elections; Osaka Mayor election in November 2011 and the election of Osaka and Tokyo prefectural constituencies of the House of Councillors of Japan in July 2010. The daily weblog posting data is monitored by using the social media monitoring platform "Kuchikomi Kakaricho" of Hottolink for these three elections. For the recent Osaka Mayor elections, the Twitter posting is also monitored. For the three elections, the exposure of election candidates on television is monitored by using the data set presented by M Data.

The data of the weblog posting and the exposure data on television are collected about one month duration before the

election day. The formal election season for Osaka Mayor election is 14 days and that for the House of Councillors is 17 days.

We show our observed data for the Osaka Mayor election. The data for the election of Osaka and Tokyo prefectural constituencies of the House of Councillors will be used in the calculated result as comparison. For Osaka Mayor election in November 2011, the two major candidates are Toru Hashimoto and Kunio Hiramatsu. Toru Hashimoto was the Governor of Osaka prefecture. He is lawyer and is very popular as an entertainer on Japanse television. Kunio Hiramatsu was the Mayor of Osaka City. Toru Hashimoto get the election.

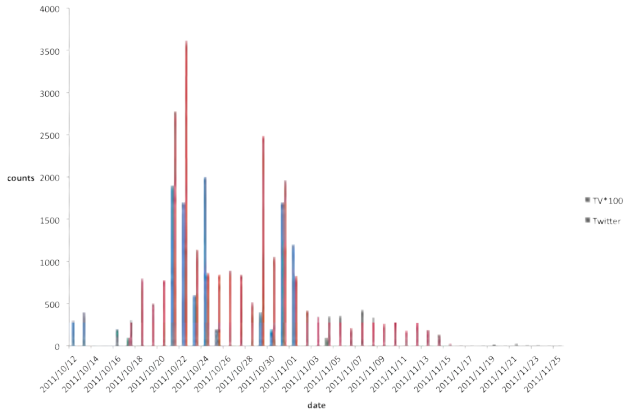

Figure 1 Daily Twitter data and TV exposure data for Kunio Hiramatsu of Osaka Mayor election. The blue is TV exposure and the red is the Twitter. The horizontal axis is the date. The last of the date corresponds to the one day before the election.

.

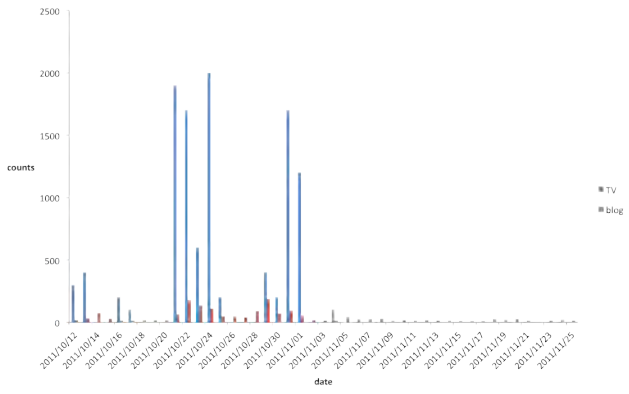

Figure 2 Daily weblog data and TV exposure data for Kunio Hiramatsu of Osaka Mayor election. The blue is TV exposure and the red is the weblog. The horizontal axis is the date. The last of the date corresponds to the one day before the election.

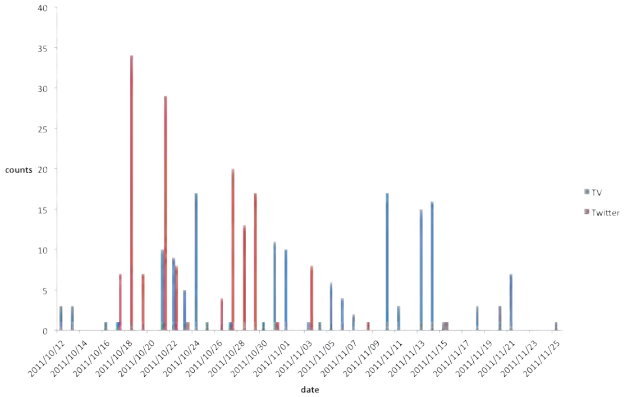

Figure 3 Daily Twitter data and TV exposure data for Toru Hashimoto of Osaka Mayor election. The blue is TV exposure and the red is the Twitter. The horizontal axis is the date. The last of the date corresponds to the one day before the election.

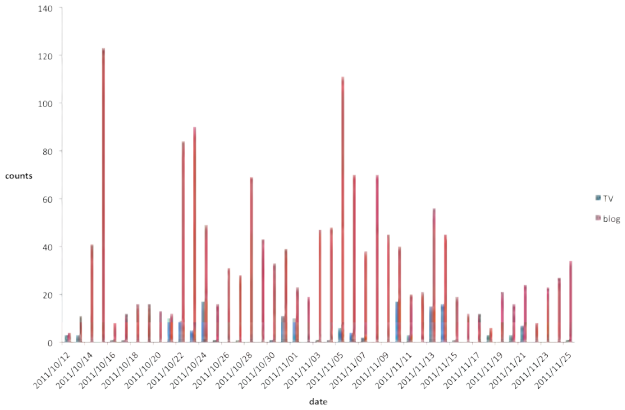

Figure 4 Daily weblog data and TV exposure data for Toru Hashimoto of Osaka Mayor election. The blue is TV exposure and the red is the weblog. The horizontal axis is the date. The last of the date corresponds to the one day before the election.

Figs.1-4, we show the observed daily weblog posting data, daily Twitter data and the daily exposure data on television in Osaka area . The horizontal axis mean the date and the last day of the horizontal axis is the last day of election season (one day before the vote).

From the data, we found that there is a lot of tweet for Toru Hashimoto. For weblog, the difference between Kunio Hiramatsu and Toru Hashimoto is very small. Especially for the weblog for Kunio Hiramatsu in fig.1, we found the strong correlation between the daily number of weblog and the daily TV exposure. Kunio Hiramatsu is not popular in Twitter as we can find in fig.2. For the weblog, we found the very stable weblog posting for Toru Hashimoto.

Mathematical Model

Vote Intention

In the mathematical model for hit phenomena [4-9], we use the equation for the purchase intention. Similarly, we consider here the vote intention $I_i(t)$ for each voter *i* as a function for time. We assume here that the vote intention for the voter *i* corresponds to the intention of the voter who want to vote to the candidate in the election.

The equation we use here is the equation for the vote intention as the effect of television exposure, direct communication and indirect communication. Each effect is explained below.

TV exposure

TV exposure of the election candidate in the news program, the new shows and the advertisement on television is the very important factor to increase the vote intention of the voter in the society. We consider the exposure effect as an external force in the equation of the vote intention as follows,

$$dI_i(t)/dt = C_{adv}A(t) + communications \quad (1)$$

where *A(t)* is the length of the exposure of the candidate on television and C_{adv} is the strength of the exposure of the candidate to affect the vote intention. Though the effect of exposure can be different for each person, we assume here, for simplicity, that the exposure affects to every persons equivalently. The effect of political news can be included in this exposure term additionally.

Communication effect

Usually, the campaign success in the political election is related to the spread by word of mouth (WOM). Such WOM is sometimes very significant effect to the election. Thus, such WOM effect should be included in the mathematical model of hit phenomena.

The WOM effect can be distinguished into the two types; WOM direct from friends and indirect as rumor. We name that the WOM effect from friends is *direct communication*, because voters obtain information directly from their friends. The voters should be distinguished into two; the voters who decide the vote and the voters who have not yet decided. The decided voter corresponds to the adaptor in marketing theory. Usually in previous WOM theories [10-13], the communications from the adaptor to non-adaptor have been considered. Here, in this paper, we include also the communication between non-decided voters. It is very significant for political elections, because many voters fix their decision to vote in the last of the election campaign duration. Let consider that the person "i" hear the information form the person "j". The probability per unit time to effect the information to the vote intention to the person "i" can be described as $D_{ij}I_j(t)$ where $I_j(t)$ is the vote intention of the person "j" and D_{ij} is the coefficient of the direct communication.

Rumor effect is named as *indirect communication* where the person hears the rumor from chats on a street, chats from the next table in restaurant, chats in the trains or finds the rumor in an Internet blogs and Twitter. Let consider that the person "i" hear the chat between the person "j" and the person "k", the strength of the effect of the chat can be described as $D_{jk}I_j(t)I_k(t)$. The probability per unit time to affect the chat to the purchase intention of the 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 $P_{ijk} = Q_{ijk}D_{jk}$.

Therefore, the direct communication is the two-body interaction and the indirect communication is the three-body interaction. Thus, including the effect of the direct communication and the indirect communication into the equation (11), we obtain the following equation for the mathematical model for hit phenomena,

$$\frac{dI_i(t)}{dt} = C_{adv}A(t) + \sum_{j\neq i}^{N} D_{ij}I_j(t) + \sum_{j}^{N}\sum_{k}^{N} P_{ijk}I_j(t)I_k(t) \quad (2)$$

Because of the term of the indirect communication, this equation is a nonlinear equation.

Mean field approximation

To solve the equation (2), we introduce here the mean field approximation for simplicity. Namely, we assume that the every person moves equally so that we can introduce the averaged value of the individual vote intention.

$$I = \frac{1}{N}\sum_j I_j(t) \quad (3)$$

Thus, the equation (12) is

$$\frac{d}{dt}\left[\frac{1}{N}\sum_i I_i(t)\right] = +A(t) + \frac{1}{N}\sum_i \sum_{j\neq i}^{N} D_{ij}I_j(t) + \frac{1}{N}\sum_i \sum_j \sum_k P_{ijk}I_j(t)I_k(t) \quad (4)$$

Introducing the number of potential voter N_p, we obtain the direct communication term from the person who do not dicide the vote as follows,

$$\frac{1}{N}\sum_i \sum_j D_{ij} I_j(t) = \frac{1}{N}\sum_i N \frac{1}{N}\sum_j D_{ij} I_j(t)$$
$$= \frac{N_p - N(t)}{N_p}\left(N_p - N(t)\right)\frac{1}{N_p - N(t)}\sum_j D^{nn} I_j(t) \quad (5)$$
$$= \frac{N_p - N(t)}{N_p}\left(N_p - N(t)\right) D^{nn} I(t)$$

where the summation is performed only for the persons who do not dicede the vote. D^{nn} is the factor of the direct communication between non-decided voters at the time t.

Similarly, we obtain the indirect communication term due to the communication between the person who do not decide the vote at the time t,

$$\frac{1}{N}\sum_i \sum_j \sum_k P_{ijk} I_j(t) I_k(t) = \frac{1}{N}\sum_i N \frac{1}{N}\sum_j N \frac{1}{N}\sum_k P_{ijk} I_j(t) I_k(t)$$
$$\Rightarrow \left(\frac{N_p - N(t)}{N_p}\right)^3 N_p^2 P^{nn} I^2 = \frac{\left(N_p - N(t)\right)^3}{N_p} P^{nn} I^2 \quad (6)$$

where P^{nn} is the factor of the indirect communication between the persons who do not decide the vote at the time t.

For the direct communication between decided and non-decided voters can be written as follows,

$$\frac{1}{N}\sum_i \sum_j D_{ij} I_j(t) = \frac{1}{N}\sum_i N \frac{1}{N}\sum_j D_{ij} I_j(t)$$
$$\Rightarrow \frac{N(t)}{N_p}\left(N_p - N(t)\right) D^{ny} I \quad (7)$$

where Dny is the factor of the direct communication between decided and non-decided voter. For the indirect communication, we obtain more two terms corresponding to the indirect communication due to the communication between decided voters and that between decided and non-decided voters as follows,

$$\frac{\left(N(t)\right)^2\left(N_p - N(t)\right)}{N_p} P^{yy} I^2$$
$$+\frac{N(t)\left(N_p - N(t)\right)^2}{N_p} P^{ny} I^2 \quad (8)$$

where P^{yy} is the factor of the indirect communication between decided voters and P^{ny} is the factor of the indirect communication due to the communication between decided and non-decided voters at the time t.

Finally, we obtain the equation of the mathematical model for hit phenomena within the mean field approximation as follows,

$$\frac{dI(t)}{dt} = C_{adv} A(t)$$
$$+\frac{\left(N_p - N(t)\right)^2}{N_p} D^{nn} I + \frac{N(t)}{N_p}\left(N_p - N(t)\right) D^{ny} I$$
$$+\frac{\left(N_p - N(t)\right)^3}{N_p} P^{nn} I^2 + \frac{\left(N(t)\right)^2\left(N_p - N(t)\right)}{N_p} P^{yy} I^2 \quad (9)$$
$$+\frac{N(t)\left(N_p - N(t)\right)^2}{N_p} P^{ny} I^2$$

where

$$N(t) = N_p \int_0^t I(\tau) d\tau \quad (10)$$

Thus, the equation (37) with (38) is the nonlinear integro-differential equation. However, since the handling data is daily, the time difference is one day, we can solve the equation numerically as a difference equation.

Calculated Results

Reliable factor

Using the equation (9) and (10), we calculate the vote intention for Japanese political elections where the TV exposure data presented from M Data is inputted into A(t) with the unit of seconds. The calculated result is compared with the daily blog entry counts observed by using the social media monitoring platform “Kuchikomi Kakaricho” of Hottolink in Japan. In order to adjust the parameters to fit the calculation with the observed blog data, the reliability of the adjustment is required.

For the purpose of the reliability, we introduce here the so-called “R-factor” (reliable factor) well-known in the field of the low energy electron diffraction (LEED) experiment [14]. In the LEED experiment, the experimentally observed curve of current vs. voltage is compared with the corresponding theoretical curve using the R-factor.

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

$$R = \frac{\sum_i (f(i) - g(i))^2}{\sum_i (f(i)^2 + g(i)^2)} \quad (11)$$

where the function f(i) and g(i) is defined in figure 5. The smaller R, the function f and g show that better matches.

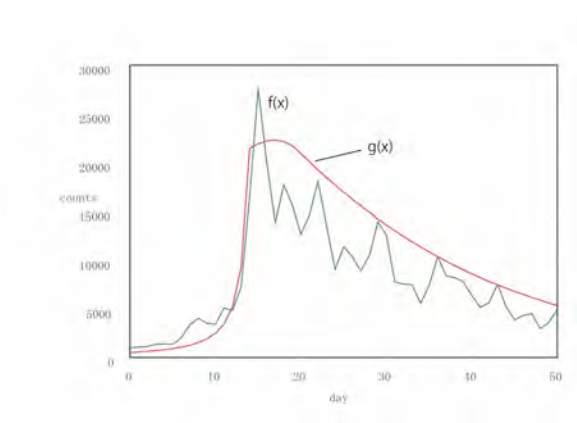

Figure 5 Example to explain the R-factor

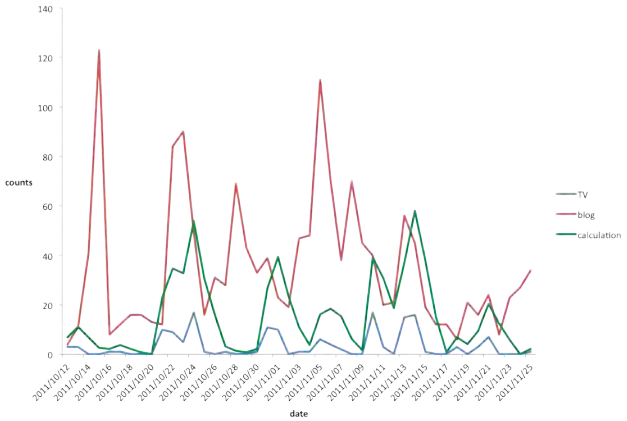

Figure 6 Calculation for Kunio Hiramatsu. The horizontal axis is the date. The last day of the axis is one day before the vote. The blue, the red and the green curve correspond to TV exposure, the weblog and our calculation.

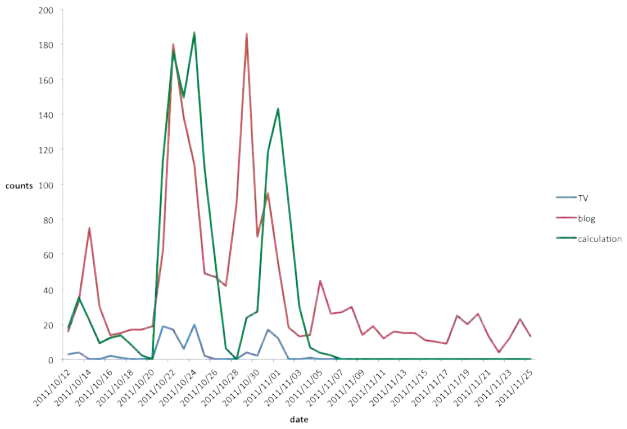

Figure 7 Calculation for Toru Hashimoto. The horizontal axis is the date. The last day of the axis is one day before the vote. The blue, the red and the green curve correspond to TV exposure, the weblog and our calculation.

Calculation for Osaka Mayor election

. The calculation of the vote intention is shown in fig.6 and 7 for Kunio Hiramatsu and Toru Hashimoto. The parameters we adjust to the daily weblog data is the strength of TV exposure C_{adv} , the strength of the direct communication D and the strength of the indirect communication P. Here we assume that $D^{nn} = D^{ny}$ and $P^{nn} = P^{ny} = P^{yy}$ for simplicity. For Kunio Hiramatsu, we adjust our calculation to the daily weblog of the last 10 days of election season. For Toru Hashimoto, we adjust our calculation to the first high peak of the daily weblog. The parameters we obtain are shown in Table 1.

	Hiramatsu	*Hashimoto*	*Hashimoto Twitter*
Cadv	2.3	6.0	115
D	0.003	0.02	0.0009
P	0.00009	0.0002	0.00002

Table 1 Parameters of our calculation.

From table 1, we found that every parameters of Toru Hashimoto are larger than that of Kunio Hiramatsu. Large parameter means large popularity. Since the Toru Hashimoto get the election, our obtained value of the strength of TV exposure, direct and indirect communication are consistent.

House of Councillors election in Osaka

For the election of Osaka and Tokyo prefectural constituencies of the House of Councillors of Japan in July 2010, there were four major candidates, Ishikawa, Kitagawa, Odate and Okabe. M Okabe is very popular female entertainer having some regular TV programs. In Table 2, we show the number of weblog posting for the four candidates during the election duration.

Candidate	*number of posting*
Okabe	4424
Odate	1254
Kitagawa	858
Ishikawa	701

Table 2 Total number of posting in House of Councillors election in Osaka, 2010.

From the table 2, we found that the entertainer Ms Okabe collect a lot of posting. However, the results of election is shown in table 3 that the order of candidate in the real vote is completely upside down. Thus, we cannot use the total number of weblog posting as the key parameters of the election.

Candidate	*number of votes obtained*
Ishikawa	864278
Kitagawa	706986
Odate	698933
Okabe	617932

Table 3 Results of the votes of House of Councillors election in Osaka, 2010.

	Ishikawa	*Kitagawa*	*Odate*	*Okabe*
Cadv	50	50	50	190
D	0.093	0.001	0.001	0.001
P	0.00217	0.00628	0.0061	0.0014

Table 4 Parameters of model for the vote of House of Councillors election in Osaka, 2010.

We perform the same calculation of the vote intention for the House of Counsillors election in Osaka. The parameters we obtain from the adjusting of our calculation to the daily weblog posting is shown in table 4. From table 4, we found that the entertainer Mari Okabe obtain large value for the strength of TV exposure. However, for the direct communication strength D, Ishikawa obtain the largest value. For Kitagawa, Odate and Okabe, the indirect communication strength values are different though the direct is same.

Therefore, as an assumption, we test the following communication strength,

$$E = D + \beta P \qquad (12)$$

the sum of the direct and the indirect communication strength. If we set $\beta = 1$ for simplicity, we obtain table 5.

Candidate	*E*
Ishikawa	0.0095
Kitagawa	0.0072
Odate	0.0071
Okabe	0.0024

Table 5 Communication strength E for the four candidates of House of Counsillors election in Osaka, 2010.

Table 5 agrees very well with table 3, the final results of the election of the House of Counsillors in Osaka 2010. It means that the communication strength E defined by eq.(12) can predict the result of election. The communication strength E for $\beta = 1$ explain the results of the Osaka Mayer election, also.

To confirm it, we perform one more example of the election, the House of Counsillors election for Tokyo prefectural constituency. The result of our calculation is shown in table 6. Table 6 show us that the communication strength E for $\beta = 1$ explain the results of the House of Counsillors election for Tokyo prefectural constituency.

	Renho	*Taketani*	*Nakagawa*	*Ogawa*
Cadv	37	6.8	5.3	5.9
D	0.09	0.06	0.008	0.003
P	0.0000037	0.005	0.009	0.003
E	0.0900037	0.065	0.017	0.006
votes	1710734	806862	711171	696672

Table 6 TV exposure strength Cadv, D, P and communication strength E for the four candidates of House of Counsillors election in Tokyo, 2010.

.

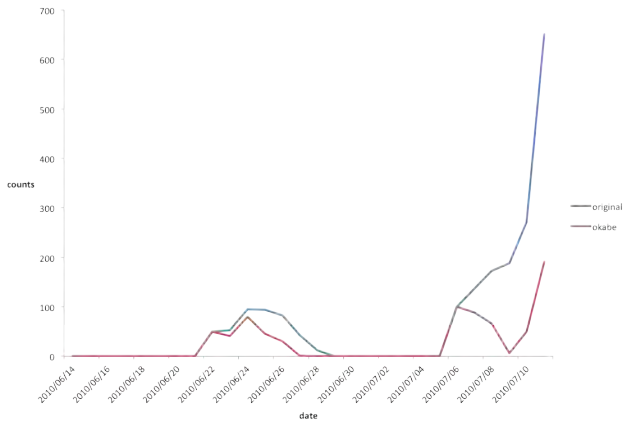

Figure 8 Calculation of the vote intention of Ishikawa using the parameter D and P of Okabe. The blue is the original and the red is the calculation using Okabe parameters of D and P.

Discussion

From our results, we found that the communication strength E defined by eq.(12) is very consistent with the real election result. For the House of Counsillors election in Osaka. our result explain why the famous entertainer Mari Okabe lost the election. Though she has very large TV exposure strength, the communication strength is very weak. It means that the posting on Mari Okabe did not continued after the TV exposure. In fig.8, we show the calculation of the vote

intention of Ishikawa using the parameter D and P of Okabe. As we can see, the vote intention is clearly different.

For the House of Counsillors election in Osaka, we found that the total number of posting has no meaning for the analysis of elections. Instead, the communication parameter E we introduced here using the mathematical model for hit phenomena is significant to predict the political election.

Important point for the actual applicaton of our study for political elections is that the communication parameter E can be obtained everytime, even before the formal duration of election. Thus, our result can be applied to decide the candidate of each constituency far before the election.

CONCLUSION

We present the mathematical model of hit phenomena as an equation for the vote intention for each voter. We found that the communication strength measured from social media using our model can be applicable for the analysis of political and can predict the result of political elections. our theory.

References

[1] Echo Keif, www.pubchoicesoc.org/paper_2012/Keif.pdf, George Mason University, 2012

[2] Zahara Harb, M/C Journal 14 (2011) #039

[3] K.Endo, J.Socio-Informatics 3 (2010) 47

[4] A.Ishii and N.Yoshida, Reports of the Faculty of Engineering Tottori University 36 (2005) 71 (in Japanese)

[5] A.Ishii, N.Yoshida, H.Arakaki and F.Yamazaki, Reports of the Faculty of Engineering Tottori University 37 (2007) 107 (in Japanese)

[6] N.Yoshida, A.Ishii and H.Arakaki, "*Daihitto no Houteishiki* (Equation of Big Hit)" Discover Twenty One, Tokyo, 2010, ISBN-10: 4887598467 (in Japanese)

[7] A Ishii, S Umemura, T Hayashi, N Matsuda, T Nakagawa, H Arakaki, N Yoshida, arXiv:1002.4460 (2010)

[8] A.Ishii, H.Arakaki, N.Matsuda, S.Umemura, T.Urushidani, N.Yamagata and N.Yoshda, to be published; arXiv:1112.1143v1 [physics.soc-ph]

[9] A.Ishii, T.Matsumoto and S Miki, Prog.Theor.Phys. in press; arXiv:1112.0767v1 [physics.soc-ph]

[10] F.Bass, Management Science 15 (5): p215–227 (1969).

[11] F. M. Bass, "The Adoption of a Marketing Model : Comments and Observations," in Innovation Diffusion Models of New Product Acceptance, V. Mahajan and Y. Wind, eds., Ballinger, 1986.

[12] C Dellarocas, NF Awad and X Zhang, 2004 working paper, MIT Sloan School of Management

[13] C Dellarocas, X Zhang and NF Awad, J.Interactive Marketing 21(4) (2007) 23-45

[14] J.B.Pendry. J.Phys.C3 937 (1980)