

Finding Influencers and Consumer Insights in the Blogosphere

Naohiro Matsumura

Osaka University

Hikaru Yamamoto

Seikei University

Daisuke Tomozawa

NIFTY Research Institute

Agenda

- Introduction
- Related work
- Influence Diffusion Model
- Empirical Analysis
- Summary and future work

Challenge for marketers

- To measure WOM in the blogosphere and utilize the results for marketing decision making
 - Who talks about the product and influence others?
 - Who is influenced by WOM?
 - What do consumers like/dislike about the product?
 - Which marketing element is perceived to be relevant to consumers?
 - Which communication messages successfully reach consumers?

Objective

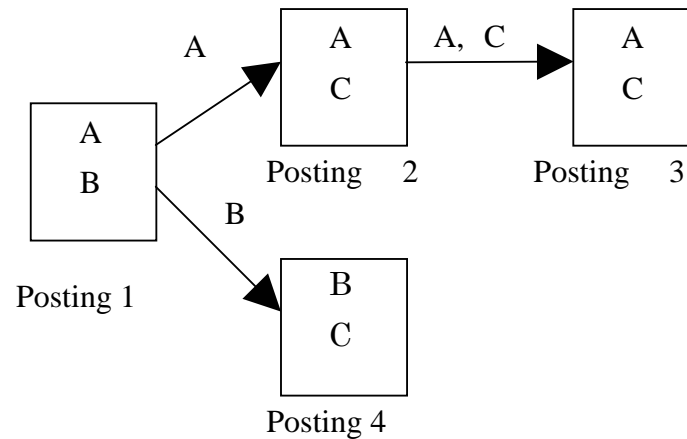
- To identify influencers who write blogs which influence other bloggers
- To identify the key marketing elements which diffuse via WOM communication in the blogosphere

Related Work

- Marketing
 - WOM as a driver of buyer behavior (Arndt 1967, Engel, Blackwell, and Kegerreis 1969, Day 1971)
 - Finding opinion leader (Lazarsfeld, Berelson, and Gaudet 1948, Childers 1986; King and Summers 1970; Rogers 1962)
 - Social structure and diffusion process of WOM (Reingen and Kernan 1986; Brown and Reingen 1987)
- CS
 - link behavior with large-scale online data and focuses on the network structure (Herring et al. 2005; Kumar, Novak, and Tomkins 2006; Marlow 2004)
 - influence of nodes in the blogosphere (Java et al. 2007; Adar and Adamic 2005; Furukawa et al. 2007)

Influence Diffusion Model

- What does it do?
 - calculates the spread of a term in the blogosphere recursively, and evaluates the influence of terms, blog entries, and bloggers



Influence of Postings

- The number of terms propagating from a posting x to a posting y

$$n_{x \rightarrow y} = |w_x \cap \dots \cap w_y| \quad (1)$$

W_x : set of terms included in posting x

$$n_{1 \rightarrow 2} = |w_1 \cap w_2| = 1$$

$$n_{1 \rightarrow 3} = |w_1 \cap w_2 \cap w_3| = 1$$

$$n_{1 \rightarrow 4} = |w_1 \cap w_4| = 1$$

$$n_{2 \rightarrow 3} = |w_2 \cap w_3| = 2$$

- The outgoing influence of posting x

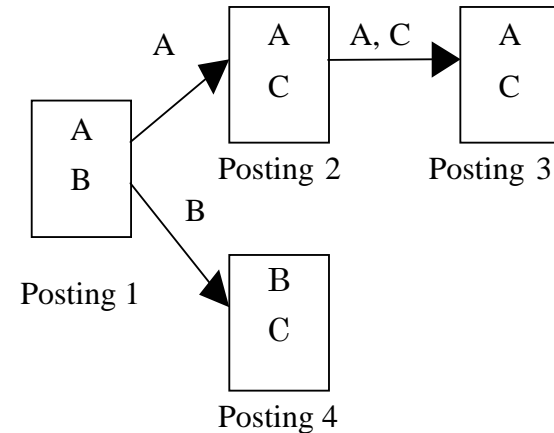
$$i_x = \sum_{y \in \text{all_postings}} n_{x \rightarrow y} \quad (2)$$

$$i_1 = n_{1 \rightarrow 2} + n_{1 \rightarrow 3} + n_{1 \rightarrow 4} = 1 + 1 + 1 = 3$$

$$i_2 = n_{2 \rightarrow 3} = 2$$

$$i_3 = 0$$

$$i_4 = 0$$



- The incoming influence of posting x

$$j_x = \sum_{y \in \text{all_postings}} n_{y \rightarrow x} \quad (3)$$

$$j_1 = 0$$

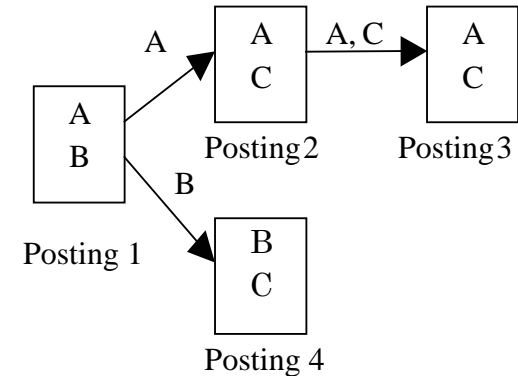
$$j_2 = n_{1 \rightarrow 2} = 1$$

$$j_3 = n_{1 \rightarrow 3} + n_{2 \rightarrow 3} = 1 + 2 = 3$$

$$j_4 = n_{1 \rightarrow 4} = 1$$

Influence of Bloggers

- Outgoing Influence of blogger x (I_x)
 - S_a : the author of posting 1
 - S_b : the author of posting 2
 - S_c : the author of posting 3 and postings 4



$$I_x = \sum_{y \in \text{all_postings_by_}x} i_y \quad (4)$$

$$I_a = i_1 = 3$$

$$I_b = i_2 = 2$$

$$I_c = i_3 + i_4 = 0$$

- Incoming influence of blogger x (J_x)

$$J_x = \sum_{y \in \text{all_postings_referred_by_}x} j_y \quad (5)$$

$$J_a = j_1 = 0$$

$$J_b = j_2 = 1$$

$$J_c = j_3 + j_4 = 3 + 1 = 4$$

Influence of Terms

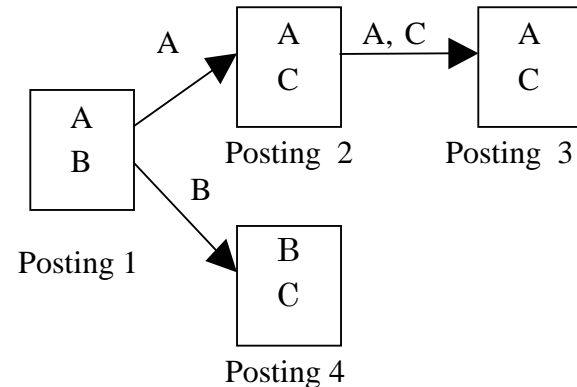
$$\delta_{x \rightarrow y}(A) = \begin{cases} 1 & \text{if } \{w_x \cap \dots \cap w_y\} \text{ contains } A \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

K_x : influence of term x

$$K_a = \delta_{1 \rightarrow 2}(A) + \delta_{1 \rightarrow 3}(A) + \delta_{2 \rightarrow 3}(A) = 3$$

$$K_b = \delta_{1 \rightarrow 4}(B) = 1$$

$$K_c = \delta_{2 \rightarrow 3}(C) = 1$$



- Only the propagated terms among link relationships are counted!
 - Captures frequent terms since they have a better chance to be propagated
 - Picks up infrequent terms that are considered to be valuable by consumers.

Empirical Analysis

- Applied IDM to blog data
 - collected by BuzzPulse (NIFTY research institute)
 - a blog mining and consulting service to advertisers
 - collects over 200 million blogs, approximately 90% of all blogs in Japan
- Topic of the analysis
 - New shampoo product “Tsubaki (camellia)”
- Data Description
 - A total of 10,864 postings
 - March to August 2006
 - blogs containing the term “Tsubaki”
 - Links
 - Total 3,744
 - Average 0.34
 - Trackbacks
 - Total 854
 - Average 0.08

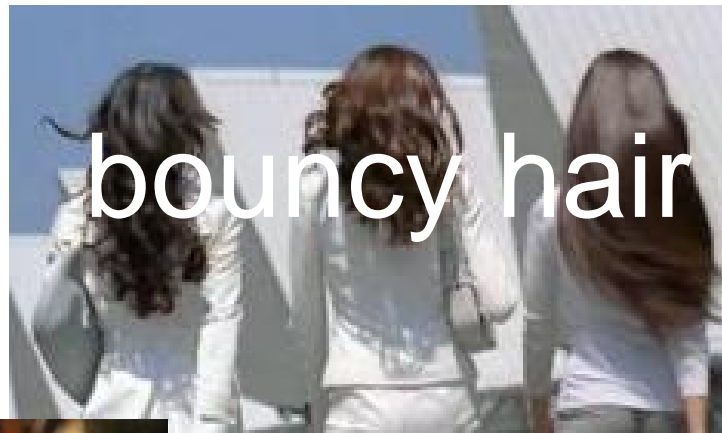


Data preparation

- Tag by the stage of consumer purchase behavior
 - classified manually!

tag	# of article
Customer	2,763
Potential Customer	3,952
Don't know	38
Cannot Open	357
Irrelevant	3,754
Total	10,863

Marketing Elements of TSUBAKI



package



Top Influential Terms

	Term	TF	Term	IDM	diff in ranking
1	Shiseido	3182	Shiseido	970	0
2	shampoo	2921	CM	532	1
3	CM	2820	feeling	375	1
4	feeling	2162	shampoo	155	-2
5	like	2035	effect	146	20
6	me	1609	SMAP	140	1
7	SMAP	1556	special	129	520
8	product	1344	classification	129	950
9	woman	1270	series	126	123
10	on sale	1221	photo	83	4
11	fragrance	1160	starring in CM	76	17
12	expect	1156	damage	72	1439
13	actress	1041	official site	70	2860
14	photo	991	coating	70	1986
15	skincare	821	friend	68	6
16	work	818	like	65	-11
17	use	791	repair	62	446
18	news	786	conditioner	58	9
19	time	775	new arrival	57	20
20	name	704	camellia oil	57	126

Influencers and Influencees

Top bloggers by outgoing influence

	Blogger	Inf(out)	Inf(in)	Posting	link
1	http://1tsu...	1020	1302	14	84
2	http://1tsu...	801	436	16	130
3	http://b003...	358	276	6	25
4	http://b001...	150	283	6	21
5	http://blog.l...	138	24	3	9
6	http://b004...	115	91	2	3
7	http://1hear...	111	194	3	26
8	http://morn...	97	0	2	15
9	http://jean...	97	178	25	49
10	http://knty...	94	52	2	6

Top bloggers by incoming influence

	Blogger	Inf (out)	Inf (in)	Posting	link
1	http://1tsu...	1020	1302	14	84
2	http://1tsu...	801	436	16	130
3	http://b001...	150	283	6	21
4	http://b003...	358	276	6	25
5	http://amebl..	10	210	2	11
6	http://1hear..	111	194	3	26
7	http://www.b	79	186	2	16
8	http://jean...	97	178	25	49
9	http://shise...	55	123	13	4
10	http://1hear..	115	91	2	3

IDM vs. Links and Trackbacks

- % of “Customer” and “Potential Customer” included in top 100 postings by Links, trackbacks and influences (IDM)

Comparison by Posting

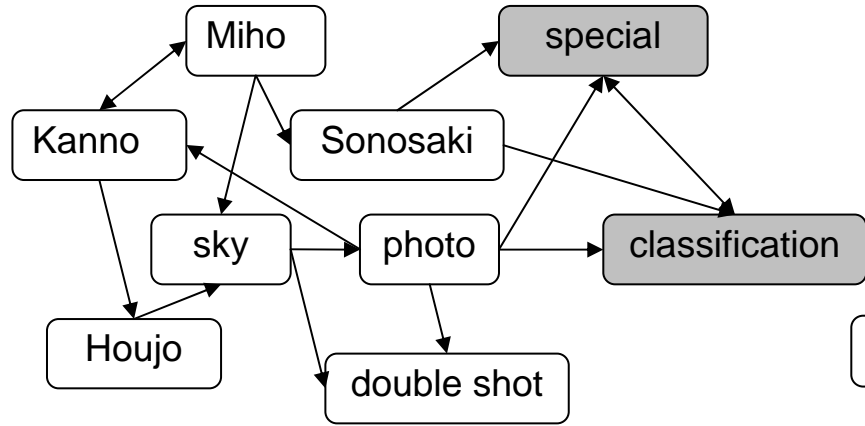
	Link	TB	Inf (out)	Inf (in)
Potential Customer	0.05	0.09	0.14	0.22
Customer	0.07	0.40	0.61	0.51
irrelevant	0.88	0.42	0.21	0.25
cannot open	0.00	0.09	0.04	0.02

Comparison by Blogger

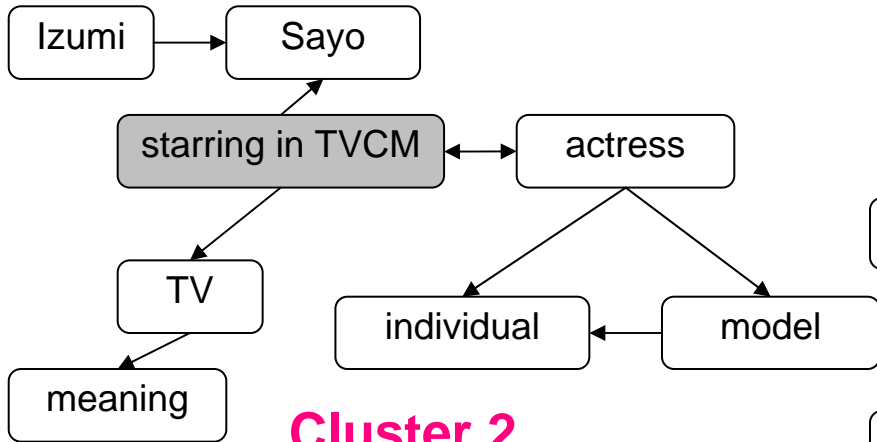
	Link	TB	Inf (out)	Inf (in)
Potential Customer	0.18	0.13	0.19	0.24
Customer	0.39	0.45	0.55	0.48
irrelevant	0.41	0.37	0.23	0.26
cannot open	0.02	0.05	0.03	0.02

IDM performs better in capturing “Customer” and “Potential Customer”

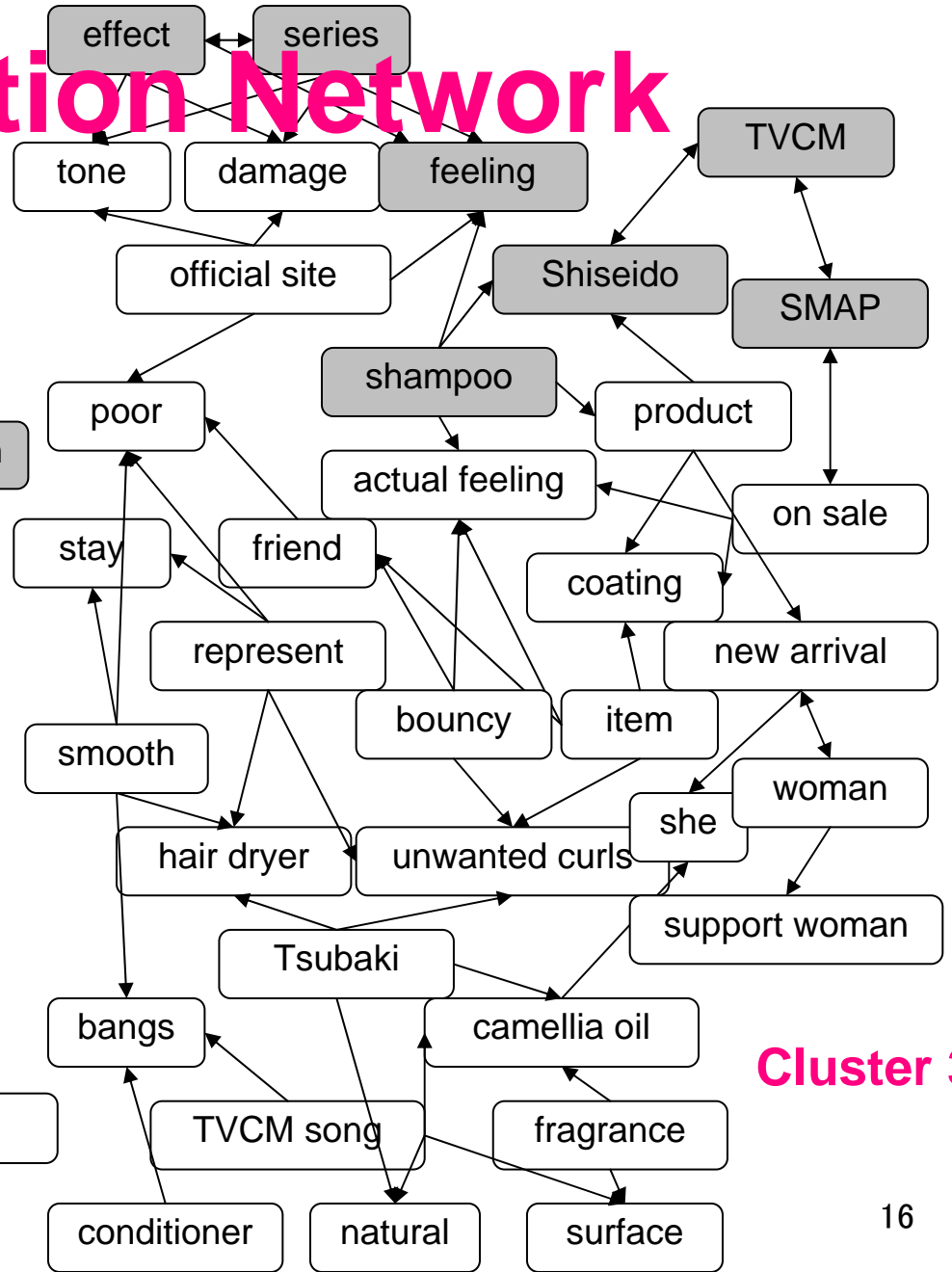
Propagation Network



Cluster 1



Cluster 2



Cluster 3

Summary

- By applying IDM to blog data, our research
 - identified influential postings, influential bloggers, and influential terms
 - Picks up both frequent and infrequent but relevant terms
 - captured marketing-relevant blogs compared to links and trackbacks
 - Picks up more blogs with purchase experience /interest
 - presented a propagation network visualizing consumer-oriented key marketing elements

Future work

- examination of the model in other categories
- evaluating the model statistically toward formalization
- prediction of WOM behavior / purchase behavior

Thank you

Naohiro Matsumura

matumura@econ.osaka-u.ac.jp

Hikaru Yamamoto

yamamoto@econ.seikei.ac.jp

Daisuke Tomozawa

tomozawa.daisuke@nifty.co.jp