

Users' Behavioral Analysis on Weblogs

Tadanobu Furukawa

Information Technology Research Institute,
Tokyo University of Science
Yamazaki 2641, Chiba 278-8510, Japan
tektf@mt.is.noda.tus.ac.jp

Yutaka Matsuo

National Institute of Advanced Industrial
Science and Technology (AIST)
Sotokanda 1-18-13, Tokyo 101-0021, Japan
y.matsuo@aist.go.jp

Tomofumi Matsuzawa and Masayuki Takeda

Dept. of Information Sciences,
Tokyo University of Science
Yamazaki 2641, Chiba 278-8510, Japan
{t-matsu, takeda}@is.noda.tus.ac.jp

Koki Uchiyama

Hottolink, Inc.
Nishigotanda 2-11-17, Tokyo 141-0031, Japan
uchi@hottolink.co.jp

Abstract

This study investigates Weblog users' behavioral data on a hosting service in Japan. The study analyzes users' behaviors, such as browsing and bookmarking, in addition to posting comments and sending trackbacks. Results of this study reveal the causes of a user's visiting and regularly browsing behaviors, especially from a social network point of view. The correlation between the strength of two users on the social network and their visiting behavior is analyzed.

Introduction

Web logs (blogs) are receiving much attention as a new medium to describe individual experiences and opinions. Blog users sometimes visit others' blogs and write comments or send trackbacks as they update their own blogs. These users' behaviors result in relationships among blogs, which create interesting data for analyses of blog communities and topic diffusion. Recently, users have come to read blogs using RSS readers and social networking services (SNSs). We can assume that a user might use social networks to seek blogs.

Our analyses use the database of a blog-hosting service in Japan called Doblog¹, which NTT Data Corp. and hottolink, Inc. provided. Using this service, users can create and update their blogs easily. They write their blogs, change the templates, and write comments after they log in. Therefore, we can monitor users' behavior in great detail, even information showing which blogs a user has browsed and when. Doblog has a special function called "bookmark" to link to favorite blogs from one's own blog. We regard the book-

mark information as representative of browser bookmarks or blogs of a user's RSS reader.

This study analyzes users' behaviors from a social network point of view. Three types of relationship are considered: comment, trackback, and bookmark. We call these relationships blog-based relationships. Two types of behavioral relationships considered in this study are: Visit and Regular Reading. We clarify how the blog-based relationship affects users' behaviors. Mainly, two questions are investigated: Do users' blog-based relationships affect their browsing behavior? Furthermore, can we predict whether users visit or read a certain blog frequently based on blog-based relation information?

Relations between two Blogs

Three types of blog-based relations, **Comment, Trackback, and Bookmark**, are defined as follows: If user U_A , who writes blog A, comments on or trackbacks to blog B, we define a relation of Comment / Trackback from A to B. If user U_A registers blog B as a bookmark in the Doblog system, we define a relation of Bookmark from A to B.

We also define a user's browsing behavior. If user U_A browses B, we call it **Visiting** behavior from A to B²; if user A has visited user B's blog regularly, which is defined in this paper as more than 30 times, we call it **Regular Reading** behavior from A to B.

Our hypothesis is that users' (indirect) relationships affect users' browsing behavior. Therefore, (i) What relations among Comment, Trackback, and Bookmark behaviors engender Visiting behavior and Regular Reading behavior?

We analyzed data of 1647 blogs with top frequent users, who are 10% of all Doblog users. The data consist of 59% of

Copyright © 2006, American Association for Artificial Intelligence (www.aaai.org). All rights reserved.

¹©NTT Data Corp., ©hottolink, Inc.,
http://www.doblog.com/, using data of Oct. 2003 – Jul. 2004

²To exclude an unintended browse, we define Visit as the relation by which a user browses more than once. Comment and Trackback relations are defined similarly.

all Bookmark data, 64% Comment data, and 64% Trackback data.

Relations among two blogs

Two blogs sometimes have numerous relations. Because two blogs' direct relations are explicit and readily apparent, we specifically examine indirect relations between two blogs. A measure of each relational strength is represented by the number of routes that connect two blogs in two hops with the same direction. For example, three routes exist between A and B, as shown in Fig. 1. We can also consider two blogs that are connected by three or more hops, but that relation should be less effective. Moreover, in sociological studies, a two-hop relation is sometimes assessed, e.g., transitivity and clustering coefficients. For that reason, we examine only the number of routes in two hops.

Figure 2 shows the extent to which the rate follows the transitivity rule in these two hops. This figure reveals a positive correlation between the number of routes and the transitivity rule.

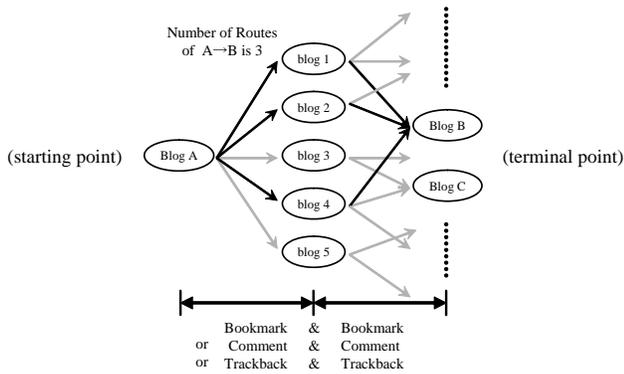


Figure 1: Number of routes

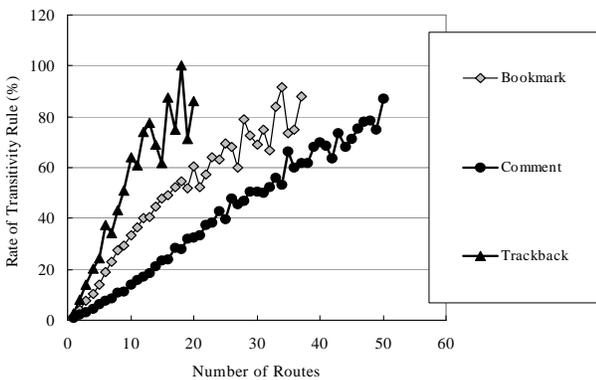


Figure 2: Number of routes and the transitivity rule

Figure 3 shows the visiting rate versus the number of routes. For example, if there are 20 routes of two-hop comment relations, the probability of that user visiting another's

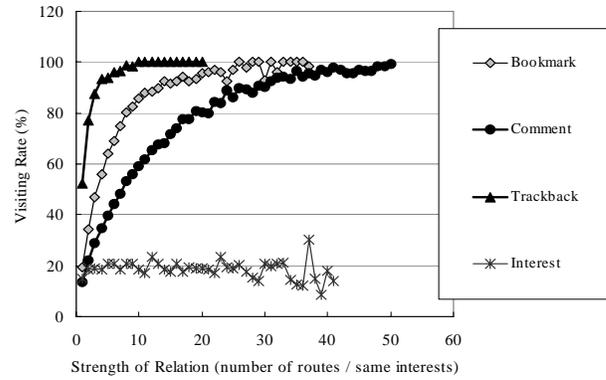


Figure 3: Number of link routes or similarity of interest and Visiting

blog is 80%. Bookmark and trackback relations more effectively induce users to visit. The reason seems to be that the Bookmarks, Comments, and Trackbacks create a hyperlink that allows a user to visit the site easily. No correlation can be found between the number of fields of identical interest and visiting behavior: Doblog allows users to register their interests categorically, which does not contribute to explain the users' behaviour.

User behavior prediction

What causes users to visit other persons' blogs? What relation most affects daily browsing behavior? If we can create a model and predict whether a user will like a blog or not, we can build a recommendation list of blogs for each user. Below, a classifier to predict a user's browsing behavior is produced using a machine learning technique.

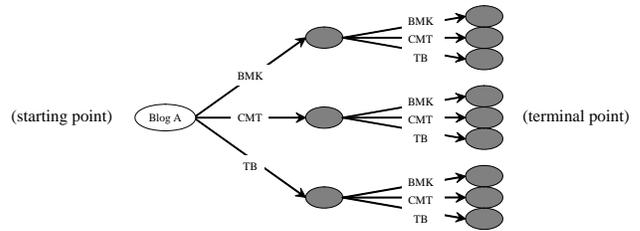


Figure 4: 12 kinds of relations. (BMK, Bookmark; CMT, Comment; TB, Trackback)

As Fig. 4 shows, 12 kinds of relation are considered in the range of two-hop relations (represented as gray nodes in the figure). For all two-blog sets (${}_{1647}P_2$ sets), we checked which relations among the 12 kinds of relations existed, and whether they have a Visiting / Regular Reading relation. Using each set as training data, we analyzed which relation determines Regular Reading behavior. Then we constructed a decision tree using a machine learning algorithm C4.5. During Visiting analysis, we excluded blog sets with a one-hop relation because there must be a Visiting behavior.

Table 1: Form of training data for analyses of relation types

2 blogs		12 Types of Relation				Vst
start	term	Bmk	Cmt	...	Tb-Tb	(Rd)
A	B	T	F	...	T	T
A	C	F	T	...	F	F
B	C	T	T	...	T	T
⋮				⋮		⋮

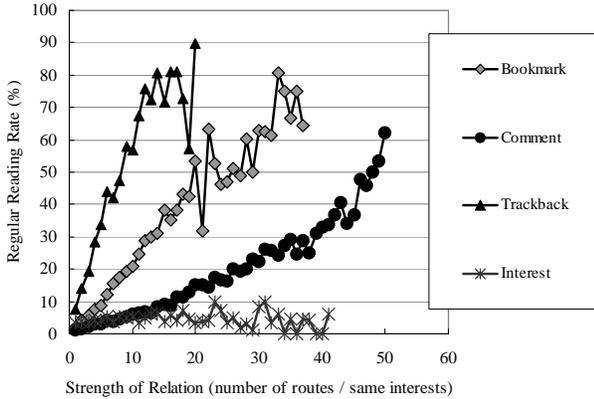


Figure 5: Number of link routes or similarity of interest & Regular Reading (more than 30 visits)

For Visiting analysis, the training data were 568,046 sets of two blogs that have more than one route by any two-hop relation. For Regular Reading analysis, 154,549 sets of two blogs have a Visiting relation. Each data set for machine learning consists of the number of routes of each relation and whether a Visiting or Regular Reading relation exists (Table 1).

Performance of the prediction method is shown in Table 2. The recall is not high, implying that it is difficult to infer all the blogs that a user visits or regularly reads because there might be numerous reasons to read blogs. However, the precision is high: if a user is in a certain (two-hop) relation to a blog, a user is likely to visit or regularly read it. Therefore, we can reasonably recommend blogs that are predicted to have a visit or regular read relation, but which have not yet been actually visited or regularly read by the user.

Figures 6 and 7 respectively show decision trees for Visit and Regular Reading. In those decision trees, the highly influential relation occupies the upper position as a node, and B / C / T represents Bookmark / Comment / Trackback. The leaves show “Yes” or “No”, representing whether the Visiting or Regular Reading relation is true or false. The

Table 2: Recall and Precision

	Recall	Precision	Error
Visit	0.152	0.633	0.310
Regular reading	0.314	0.666	0.059

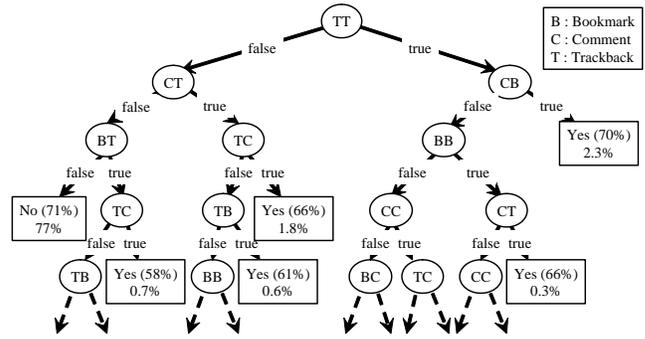


Figure 6: Decision tree of Visiting by 12 kinds of relations

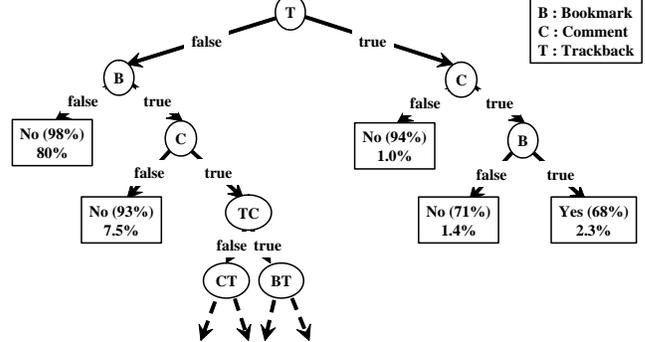


Figure 7: Decision tree of Regular Reading by 19 kinds of relations

Bookmark relation has a strong effect, but Comment is also a good index for prediction. The Comment relation implies a similarity of interests. Therefore, it is used for prediction.

Conclusions

This study explored unique relations among users’ behaviors on a blog network in which blog users visit and regularly read the materials. We analyzed the relation between the network and user behaviors, and also tried to predict whether users visit or read a certain blog frequently based on the relation information. We consider that the further analysis can form the basis of an recommendation service of blogs.

References

- Adar, E.; and Adamic, L.A. 2004. Tracking Information Epidemics in Blogspace. Workshop on the Weblogging Ecosystem, 14th International World Wide Web Conference.
- Glance, N.; Hurst, M.; and Tomokiyo, T. 2004. BlogPulse: Automated Trend Discovery for Weblogs. Workshop on the Weblogging Ecosystem, 13th International World Wide Web Conference.
- Kumar, R.; Novak, J.; Raghavan, P.; and Tomkins, A. 2003. On the Bursty Evolution of Blogspace. The 10th International World Wide Web Conference.