The Credibility of the Posted Information in a Recommendation System Based on a Map

Koji Yamamoto
Tokyo Institute of Technology
4259 Nagatsuta, Midori-ku
Yokohama, Japan
yamamoto@ntt.dis.titech.ac.jp

Daisuke Katagami
Tokyo Institute of Technology
4259 Nagatsuta, Midori-ku
Yokohama, Japan

Katsumi Nitta
Tokyo Institute of Technology
4259 Nagatsuta, Midori-ku
Yokohama, Japan

Akira Aiba
Shibaura Institute of Technology
307 Fukasaku, Minuma-ku
Saitama, Japan

Hitoshi Kuwata
Shibaura Institute of Technology
307 Fukasaku, Minuma-ku
Saitama, Japan

ABSTRACT

We propose a method for estimating the credibility of the posted information from users. The system displays these information on the map. Since posted information can include subjective information from various perspectives, we can't trust all of the postings as they are. We propose and integrate factors of the user's geographic posting tendency and votes by other users.

Categories and Subject Descriptors

H.4.2 [**Information Systems Applications**]: Decision support

General Terms

Algorithms, Experimentation, Human Factors

Keywords

credibility, posting, GIS, navigation, recommendation

1. INTRODUCTION

We have developed an information recommendation system which updates its information with postings and navigates user by using them. The problem of the system using posted information is that we can't trust all of the postings as they are because posted information can include subjective information from various perspectives. Generally speaking, information which has high credibility is posted by the users who have much knowledge about certain areas. The user's expertise was estimated based on the credibility of past posting information in existing methods. However, even if the knowledge of a specific genre can be estimated using existing methods, knowledge in the region can't be estimated. Our goal is to develop an information recommendation system using posted information with the method for estimating the credibility of posted information based on their regional characteristics.

2. PROCEDURE TO ESTIMATE THE CREDIBILITY

Our system assigns initial credibility to posted information if it is the user's first posting, as $R_i(g,x,y) = R_{default}$. $R_{default}$ is a default value determined in advance. If a new posting, which is located at (x_p, y_p) and whose genre is g, is posted, then $R_i(g, x_p, y_p)$ is determined by distance between (x_p, y_p) and the location (x_q, y_q) of past postings as the formula (1) and (2). P_i is the set of posted information by user i.

$$R_i(g, x_p, y_p) = \frac{\sum_{q \in P_i/p} w_{pq} * R_i(g, x_q, y_q)}{\sum_{q \in P_i/p} w_{pq}} \quad (1)$$

$$w_{pq} = \frac{1}{dist\{(x_p, y_p), (x_q, y_q)\}} \quad (2)$$

Formula (1) indicates weighted average of distance between the user's past postings and new posting. Function $dist$ indicates distance between two postings. Thus, posting has influence to the same user's vicinal postings. This process gives new information initial credibility. If the credibility of user's posting is relative high, new posting also has high initial credibility.

After a user browses posted information, he can vote for it. When the user votes, he chooses his rating for posting (helpful, moderate, not helpful). In case of "helpful", credibility of the voted information increases.

When user j votes for user i's information p, the system updates credibility of information p by following formula (3).

$$R_i(g, x_p, y_p) \leftarrow R_i(g, x_p, y_p) + f'(f^{-1}(R_i(g, x_p, y_p))) * V_j * \sum_{I_j \ni q} \left(R_j(g, x_q, y_q) * \exp\left(-\frac{(dist\{(x_p, y_p), (x_q, y_q)\})^2}{2{\sigma_1}^2}\right) \right) \quad (3)$$

Where V_j is 1 when vote is "helpful", or -1 when vote is "not helpful". σ_1 is a parameter adjusting influence by distance. Magnitude of the effect of vote by voter j is expressed by product of distance damping function between voter's existing posting and voted posting and credibility of voter's existing posting. Therefore, when voter's high credibility postings exist near the voted posting, the effect by such vote increases. When vote is "not helpful", V_j is negative and decreases updated value.

$f(x)$ is sigmoid function.

$$f(x) = \frac{1}{1 + e^{-\alpha x}} \quad (4)$$

WWW 2006, May 23–26, 2006, Edinburgh, Scotland.
ACM 1-59593-332-9/06/0005.

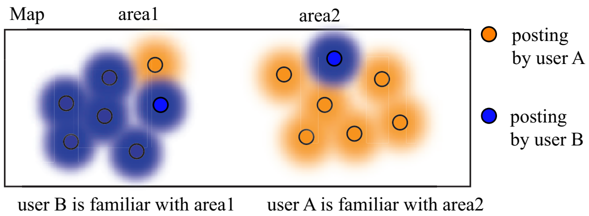

Figure 1: Geographical posting tendency: Scattered circles denote posting on a map.

This function is used to control increment of gradient when $R_i(g, x_p, y_p)$ is updated. Gradient is obtained using $f^{-1}(x)$, when value of $R_i(g, x_p, y_p)$ is equal to $f(x)$. α is the gain of sigmoid function. Increasing this value, gradient becomes steeply.

We defined geo_{ip}, which effects credibility of user i' s posting p, based on user i's geographical posting tendency.

$$geo_{ip} = f'(f^{-1}(R_i(g, x_p, y_p)))* \sum_{(P_i/p) \ni q} \left(R_i(g, x_q, y_q) * \exp\left(-\frac{(dist\{(x_p, y_p), (x_q, y_q)\})^2}{2{\sigma_2}^2}\right)\right) \quad (5)$$

σ_2 is a parameter which adjusts effect by distance between the same user's postings. This effect geo_{ip} rises when much high credibility information are located near the information p. We will call geo_{ip} "*posting tendency effect*"(Fig. 1). In fact, this effect expresses the fact that a posting is more credible by the same user posted good information near it. This revision by formula (5) is performed when someone posts or votes (formula (1) or (3)). Finally, we defined revised credibility of information p by user i as I_{ip}.

$$I_{ip} = R_i(g, x_p, y_p) + R_i(g, x_p, y_p) * geo_{ip} \quad (6)$$

geo_{ip} is added after multiplied by $R_i(g, x_p, y_p)$ because we emphasize credibility of that point.

Procedure of our method is summarized as follows. We define either new posting or vote as one step.

```
At each step(either new posting or vote is done):
If new information is posted {
1. Determine initial credibility of new posting (1)
2. Update posting tendency effect about the posting
user's all postings (5)
3. Update credibility about the posting user's
all postings (6)}

Else if a posting is voted {
1. Update credibility of voted posting (3)
2. Update posting tendency effect about the voted
user's all postings (5)
3. Update credibility about the voted user's
all postings (6)}
```

3. EXPERIMENT

To confirm our model can estimate credibility, We conducted following experiment by using the system. At first, we gathered posted information from users as preparation for experiment. We asked 20 students to register their profile, and to post the information about 4 areas around Tokyo (Shibuya, O-okayama, Machida, Aobadai), Japan, using our system. We asked them not only to post the information but also to browse and vote for other user's information. We set parameters $R_{default} = 0.2$, $\alpha = 0.2$, and considering scale of 4 areas, we used $\sigma_1 = 30(km)$ and $\sigma_2 = 15(km)$ at distance function. As a result, 134 posts and 412 votes are collected. Among them, 179 votes are "helpful", 193 votes are "moderate", and 40 are "not helpful".

Table 1: Coefficient of correlation between the average rating of subjects and model calculation

	Coorelation coefficient	Rank correlation coefficient
Case 1	0.430	0.403
Case 2	0.509	0.549
Case 3	0.731	0.780

As the next step, we asked 15 students to assign the rate of credibility to each posted information from rating scale of 1(unreliable) to 7(reliable). We performed the rating experiments for the following three cases.

- Case 1: Only posted information is displayed. Posting user's ID of information is invisible.
- Case 2: In addition to posted information, posting user's ID is displayed. Therefore subjects can figure out who posted each information.
- Case 3: In addition to the posted information and posting user's ID, position of posted information on map is displayed. Subjects can figure out not only who posted information but also where the user posted. In addition, voting user's IDs are disclosed.

In all cases, contents of information are the same.

Therefore, if subject's rating differs by cases, and if credibility calculated by our model can be approximate to rating by human at case 3, then we can say our system determines credibility as substitute for users.

Table 1 shows results. We examined correlation coefficient and rank correlation coefficient between subject's rating and credibility which was calculated by our method. Both of them, correlation of case 3 is the highest. Consequently, we consider that subjects imagined what kind of person is the posting user from his posting, and their rating got closer to the model. In conclusion, rating by subjects is different according to the situation, and this model is effective and able to consider the change in human psychology, and to calculate close credibility to rating by human.

4. CONCLUSIONS

We have proposed a method for estimating credibility of posted information on a map. From experimental results, we confirmed that our method can calculate credibility approximately to the rating by human. Our method can consider geographical posting tendency, which hasn't been considered in existing works about credibility on the Web.

Having generality not depending on contents of information, our method is applicable to various communities.

5. REFERENCES

[1] Michael Pazzani: A Framework for Collaborative, Content-Based and Demographic Filtering. *Artificial Intelligence Review*, pp. 393-408, 1999.

[2] M. G. Vozalis and K. G. Margaritis: Collaborative Filtering enhanced by Demographic Correlation", *AIAI Symposium on Professional Practice in AI, of the 18th World Computer Congress*, 2004.

[3] G. Zacharia, A. Moukas, P. Maes: Collaborative Reputation Mechanisms in Electronic Marketplaces, *Proceedings of the 32nd Hawaii International Conference on System Sciences*, 1999.