Investigations of Interests that are Induced by Remarkers and their Remarks for Item Advertisements Based on Influencer's Recommendation

ARASAWA Komei, HATTORI Shun, KUDO Yasuo

2018 Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS)

URL: http://hdl.handle.net/10258/00010112
doi: info:doi/10.1109/SCIS-ISIS.2018.00131
Investigations of Interests that are Induced by Remarkers and their Remarks for Item Advertisements Based on Influencer’s Recommendation

Komei Arasawa  
Graduate School of Engineering  
Muroran Institute of Technology  
27–1 Mizumoto-cho, Muroran,  
Hokkaido 050–8585, Japan  
Email: 18096001@mmm.muroran-it.ac.jp

Shun Hattori  
College of Information and Systems  
Muroran Institute of Technology  
27–1 Mizumoto-cho, Muroran,  
Hokkaido 050–8585, Japan  
Email: hattori@csse.muroran-it.ac.jp

Yasuo Kudo  
College of Information and Systems  
Muroran Institute of Technology  
27–1 Mizumoto-cho, Muroran,  
Hokkaido 050–8585, Japan  
Email: kudo@csse.muroran-it.ac.jp

Abstract—In order to sell items of a genre that each user is not interested in, their latent interests of the genre have to be induced. Therefore, we focus on “Influencer,” who is the person that has a large impact on a user’s behaviors, and research a recommender system which advertises some items by utilizing influencer’s remarks via social media. This paper investigates what kinds of remarks and remarkers would induce users’ interests. As the result, we have revealed many findings. One is that “One of the factors that would induce the users’ positive interests on the movie, is the positive reputation for the movie that is included in tweets.” The other is that “If the user does not like the remarker, the user’s negative interests for the items in the remarks of the person would be induced, even though the person has famousness.”

Keywords—recommendation; influencer; inducing interests;

I. INTRODUCTION

Recommender systems can learn an user’s needs, and advertise some items and services which fulfill them, for the user. For instance, a user is interested in a genre of fashion, on the other hand, the user is not interested in a genre of movies. Then, the typical recommender systems learn that the user is interested in the genre of fashion, and advertise some fashion items which she would be delighted in. That is to say, because recommender systems rarely advertise a movie item, it is difficult for movie industries to make profits. Therefore, we research a recommender system which advertises some items of the genre that the target user is not interested in, and we aim to some contributes that not only sellers can gain new customers but also the target users can discover many unknown items. In the above-mentioned instance, if the proposed system can advertise some movie items, and induce the user’s latent interest for the genre of movies, movie industries probably gain the user as a new customer, and the user probably discovers some unknown movie items.

In the above-mentioned instance, a recommender system usually needs the user’s preferences on movies to advertise some movie items for the user (e.g., the user likes mystery movies, and does not like horror movies). However, because the user is not interested in the genre of movies and there is a possible that the user’s behavior histories (e.g., purchase histories, retrieval histories) on movies are not enough, the system may not be able to gain her preferences on movies. Therefore, in recent years, “Cross-domain recommendation” [1] which advertises a domain (genre) which the system wants to recommend, by using information on other domains, has been studied. For instance, cross-domain recommender systems can advertise a mystery novel for a user who is interested in some mystery movies, by using a common feature such as “mystery.” And it means the domain of novel and the domain of movies are crossed, for advertising some novel items by using information on the user’s preferences on movies. In the above-mentioned instance, can the system advertise some movie items by using her preferences on fashion? We guess that it is difficult for the system to utilize her preferences on fashion domain when it advertises a movie item, because there is not a common feature between movie domain and fashion domain, that is to say, the genre the user is interested in would not be similar to the genre the user is not interested in, at all. Thus, there is a problem that recommender systems are hard to estimate items which the user would be delight in, from items of the genre the user is not interested in.

Therefore, as one of the solutions for the problem, we focus on “Influencer,” who is the person that has a large impact on a user’s behaviors. Figure 1 shows an overview of our proposed system. Let’s assume that a fashion model is one of the influencers for the user in the above-mentioned instance, and also the influencer remarked via social media that “Harry Potter was impressive.” Then, if the user finds the remark, the user might think that “I also want to watch some movies of Harry Potter, because my favorite person introduces them,” that is to say, the user might become interested in the movie domain. Such a system which advertises some items of the genre that the user is not interested in, by utilizing influencer’s remarks via social media, is called as “Item Advertisements Based on Influencer’s Recommendation” in our research.
III. HYPOTHESES ON INDUCING INTERESTS

This paper regards “movie” as the target domain of the advertisement. And this paper sets up 3 hypotheses on the case that users’ interests of movies would be induced, when they look at a remark of someone via social media. Here, the all hypotheses have a premise that “the all remarks that users look at include a movie title.”

- **Hypothesis 1** is that a remark has not only a movie title but also the positive reputation for the movie, and the more positive the reputation is, the more likely users’ interests on the movie are induced. One of scenarios to induce users’ interests on the movie, is that “A user feels she wants to watch also other movies (e.g., same series, same genre) because she feels that a movie she watched by chance is impressive.” Further, for a movie to be cue of the scenario, the movie has to have a factor that the user feels it is impressive. Therefore, we guess that the remarks that introduce a movie positively, are likely to lead users into the above-mentioned scenario.

- **Hypothesis 2** is that the more recent the topic of the remark is, the more likely users’ interests on the movie are induced. One of scenarios to induce users’ interests on the movie, is that “A user feels she want to watch the movie, because many people around her watch a movie.” Let’s assume that the remarks that include a specific movie title can be divided into 2 types, some are remarks that introduce the movie that is showing now, and the others are remarks that introduce the movie that is not showing now. Then we guess that the remarks that introduce the movie that is showing now, are likely to lead users into the above-mentioned scenario.

- **Hypothesis 3** is that whether a user’s interests are induced, would depend on not only the contents of the remark but also its remarker. Users have complex impression to a real person (e.g., a celebrity). If a user feels a sense of closeness for a celebrity, we expect that the remarks of the celebrity are convincing for the user. And if the other user does not like the celebrity, we also expect that the remarks of the celebrity are not convincing for the user. And if the other user does not know about the celebrity, we also expect that the user will not listen to the remarks of the celebrity. Therefore, we guess that the impression for remarkers is depending on who the listener is and who the remarkers are.
IV. MATERIALS AND METHOD

A. Materials

Because the verifications of the all hypotheses need some remarks, this paper utilizes the tweets of some celebrities. Firstly, this paper extracts the accounts of some celebrities from the web site “TwiNavi” [5]. This web site organizes the accounts of many celebrities by being divided into 11 categories (e.g., Actor, Athlete etc.), and has the ranking by the number of followers in each category. Here, the 2,204 accounts that are included in the ranking by the number of followers in each category, were extracted, on 11th July 2018. Secondly, this paper extracts the tweets that are remarked by each account, from the web site “TwiLog” [6]. This web site organizes the latest 100 tweets that are remarked by each extracted account. Here, by removing replies to someone, links to website, and images, finally 124,181 tweets were extracted, on 11th July 2018.

B. Method

We employ user’s subjective evaluation, as the method for the verification of each hypothesis, and show the procedure.

1) We collect some tweets which are on the topic of movie, and divide these tweets into 5 types. And we give some subject persons 11 tweets, while hiding the type and movie title. Type 1 has 2 tweets (in 11 tweets) that “includes a movie title.” Type 2 has 2 tweets (in 11 tweets) that “includes not only a movie title but also a positive reputation for the movie.” Type 3 has 2 tweets (in 11 tweets) that “includes a movie title, and its posted time is more recently.” Type 4 has 2 tweets (in 11 tweets) that “includes a movie title and a positive reputation for the movie, and its posted time is more recently.” The tweets of Type 1 to 4 are extracted automatically, based on our proposed criteria that are explained in Section V. And Type 5 has 3 tweets (in 11 tweets) that “introduces a movie item positively,” in which are prepared manually. 5 types of tweets are compared in Table I. “Auto” in Table I shows whether the tweets of the type are extracted automatically. “Movie” in Table I shows whether the tweets of the type are on the topic of movies. “Rep” in Table I shows whether the tweets of type has the positive reputation for the movie. “Time” in Table I shows whether the posted time of the tweets of type is new.

2) The subject persons look at only the content of each tweet, and select the interest for the hidden movie title that is included in each tweet, from the some options. In addition, the subject persons select also the impression for each tweet from some other options. By comparing the differences of the interest for each movie that is included in the tweets between “Type 1 and Type 2” or “Type 3 and Type 4,” we are able to verify the Hypothesis 1. And by comparing the differences of the impression for the tweets between “Type 1 and Type 2” or “Type 3 and Type 4,” we analyze its factors. By comparing the differences of the interest for the each movie that is included in the tweets between “Type 1 and Type 2” or “Type 3 and Type 4,” we are able to verify the Hypothesis 2. And by comparing the differences of the impression for the tweets between “Type 1 and Type 3” or “Type 2 and Type 4,” we analyze its factors.

3) We give the subject persons 3 tweets of Type 5 and each posted person, from above-shown 11 tweets. The subject persons look at the content and the posted person of each tweet, and select again the interest for the hidden movie title that is included in each tweet from the some options. In addition, the subject persons select also the impression for each posted person from some other options. By comparing the differences of the interest for the movie in the tweet, between after and before giving the posted person’s name, we are able to verify the Hypothesis 3. And by using the impression for each posted person of the tweets, we analyze its factors.

V. AUTOMATICAL CLASSIFICATION OF TWEETS

This section explains the method for extracting some tweets of Type 1 to 4 automatically from the all tweets $t_i$ $(i = 1, 2, ..., T)$, based on our proposed criteria. The criteria is based on 3 factors which are shown in Table II. First factor is “Connectedness with Item,” and has a role which searches for some tweets which are on the topic of movie, that is to say, distinguishes Type 1 to 4 and others. Second factor is “Reputation for Item,” and has a role which calculates the reputation for the movie that is included in the tweet, that is to say, distinguishes “Type 1 and Type 2” or “Type 3 and Type 4.” Third factor is “Newness of Tweet,” and has a role which calculates the newness of the tweet, that is to say, distinguishes “Type 1 and Type 3” or “Type 2 and Type 4.”
A. Connectedness with Item

Because the verification of the hypotheses needs a method which judges that a tweet $t_i$ includes a movie title, we define a function $CI(t_i)$, which shows the tweet's Connectedness with an Item (movie), that is to say, whether there is a movie title in the tweet $t_i$. The function $CI(t_i)$ has the value 0.0 or 1.0. If the tweet $t_i$ includes a movie title, the function $CI(t_i)$ has the value 1.0. On the other hand, if the tweet $t_i$ does not include a movie title, the function $CI(t_i)$ has the value 0.0.

To judge that whether there is a movie title in a tweet $t_i$, this paper employs Pattern Matching between the tweet $t_i$ and the pattern of a expression that includes a movie title. In Japanese, the expression “「xxx」” is often used, when xxx is a speech by someone, or the title of a work, or a phrase that the writer wants to emphasize. By using the expression “「xxx」” and the word “映画” (that means “movie”), the expression “映画「xxx」” (that means “the movie of xxx”) is also often used. Therefore, when the system detects the expression “映画「xxx」” in a tweet $t_i$, the system recognizes that the tweet $t_i$ includes a movie title xxx, and it gives the value 1.0 to the function $CI(t_i)$, as the following formula.

$$ CI(t_i) = \begin{cases} 
1.0 & (t_i \text{ includes a movie title}) \\
0.0 & (\text{otherwise}) 
\end{cases} \quad (1) $$

B. Reputation for Item

Because the verification of the hypotheses needs a method which judges that a tweet $t_i$ includes positive reputation for a movie, we define a function $RI(t_i)$, which shows the Reputation for an Item (movie), that is to say, whether the tweet’s posted person comments a positive impression to the movie. Here, only if the tweet $t_i$ includes a movie title and the value of the function $CI(t_i)$ is 1.0, this function is calculated. The function $RI(t_i)$ has a value from $-1.0$ to $1.0$. The more positive the comments of the tweet’s posted person is, the closer to the 1.0 the value of the function $RI(t_i)$ gets. On the other hand, the more negative the comments of the tweet’s posted person is, the closer to $-1.0$ the value of the function $RI(t_i)$ gets. And when the value of the function $RI(t_i)$ gets close to 0.0, it means the reputation that is written in the tweet $t_i$, for the movie, is neutral.

To calculate the reputation that is written in the tweet $t_i$, for the movie, this paper employs a method based on the distance between the movie title and the positive (negative) words that is written in the tweet $t_i$. Firstly, the system initializes the variable $p_i$ to the value 0.0, and it represents the positiveness that is written in the tweet $t_i$, for the movie. And the system initializes also the variable $n_i$ to the value 0.0, and it represents the negativity that is written in the tweet $t_i$, for the movie. Secondly, the system searches for the positive word(s) and the negative word(s) that are shown in Table III, in the tweet $t_i$. Here, the start position of searching for the positive (negative) word(s), is the behind of the pattern-matched movie title.

And this paper defines the sentences in the tweet $t_i$ as $s^i_j$ ($j = 1, 2, ..., J$). In the tweet $t_i$, if there is a movie title in the sentence $s^i_j$, and the system detects the positive (negative) word in the sentence $s^i_{j+k}$, the variable $p_i$ ($n_i$) is updated to the variable $p_i^{new}$ ($n_i^{new}$) by the following formulas with weight $w (< 1)$, for every detecting the positive/negative word. Here, these formulas mean “the farther the distance between the sentence $s^i_j$ and the sentence $s^i_{j+k}$ is, the lower the positiveness/negativeness of the word is.”

$$ p_i^{new} = p_i + (1.0 - p_i) \times w^k \quad (2) $$
$$ n_i^{new} = n_i + (1.0 - n_i) \times w^k \quad (3) $$

Finally, the function $RI(t_i)$ of the tweet $t_i$ is calculated by the following formula. Here, this paper sets the value of the weight $w = 0.8$.

$$ RI(t_i) = p_i^{new} - n_i^{new} \quad (4) $$

C. Newness of Tweet

Because the verification of the hypotheses needs a method which judges the newness of a tweet $t_i$, we define a function $NT(t_i)$, which shows the Newness for the Tweet $t_i$. The function $NT(t_i)$ has a value from the default value $D$ ($0.0 \leq D \leq 1.0$) to 1.0. The more recent the tweet $t_i$ is, the closer to 1.0 the value of the function $NT(t_i)$ gets. On the other hand, the older the tweet $t_i$ is, the closer to the default value $D$ the value of the function $NT(t_i)$ gets.

To calculate the newness of the tweet $t_i$, firstly, the system calculates the tolerance $d$ with the number of the tweets $T_{m}$ that include at least a movie title, by the following formula.

$$ d = \frac{1.0 - D}{T_{m} - 1} \quad (5) $$

Secondly, the system rearranges all tweets based on its posted time, and calculates the order $r_i$ of a tweet $t_i$. Here, $r_i = 1$ means that the posted time of the tweet $t_i$ is the earliest. Finally, the system calculates the $NT(t_i)$ of the tweet $t_i$ by the following formula. Here, this paper sets the value of the default value $D = 0.1$.

$$ NT(t_i) = D + d \times (T_{m} - r_i) \quad (6) $$

<table>
<thead>
<tr>
<th>Positive words</th>
<th>Negative words</th>
</tr>
</thead>
<tbody>
<tr>
<td>最高 / awesome</td>
<td>最低 / terrible</td>
</tr>
<tr>
<td>良い / good</td>
<td>悪い / bad</td>
</tr>
<tr>
<td>好き / like</td>
<td>嫌い / hate</td>
</tr>
<tr>
<td>順番 / interesting</td>
<td>つまらない / boring</td>
</tr>
<tr>
<td>楽しい / fun</td>
<td></td>
</tr>
</tbody>
</table>
D. Tweet Selection for each Type

This subsection explains a method to select 2 tweets for each type (Type 1 to Type 4) from the all tweets \( t_i \), for the verification of each hypothesis. Firstly, the value of the function RI(\( t_i \)) is normalized to the range \([0, 1]\) by using a function normalize(\( \text{RI}(t_i) \)) by the following formula.

\[
\text{normalize}(\text{RI}(t_i)) = \frac{\text{RI}(t_i) - (-1.0)}{1.0 - (-1.0)}
\]

(7)

Secondly, we calculate a function score(\( t_i \)) for a tweet \( t_i \) by the following formula. Here, the value of the function score(\( t_i \)) of the tweet \( t_i \) whose value of the function CI(\( t_i \)) equals to 0.0, dose not be calculated.

\[
\text{score}(t_i) = \text{CI}(t_i) \times \text{normalize}(\text{RI}(t_i)) \times \text{NT}(t_i)
\]

(8)

Thirdly, the value of the function score(\( t_i \)) for the tweet \( t_i \) are sorted in descending order. Finally, 2 tweets for each type (Type 1 to Type 4) are selected based on Table IV.

VI. RESULTS AND DISCUSSION

The experiment was conducted on 20th July 2018. The subject persons belong to Muroran Institute of Technology and consist of 42 men and 6 women. 9 out of 48 subject persons have a great interest on movies. 26 out of the remaining have a little interest on movies. And the remaining does not have an interest on movies.

A. Verification of Hypothesis 1

As a result, this paper reveals that the positive reputation for the movie in a tweet, can induce users’ interests positively. Figure 2 shows the rate of the subject persons who select each kind of interests for the movie, depending on the types of the tweets they look at. Focusing on the explanation of each option of interests in Figure 2, we can judge that “the option (1) and the option (2) are selected when subject persons’ interests are induced positively.” And Figure 3 shows the rate of those who select each kind of impressions for the tweet, depending on the types of the tweets they look at.

Firstly, focusing on the option (2) in Figure 2, the rate of the subject persons who feel that the movie in a tweet of Type 2 would be interesting, is higher than Type 1 by 0.17 (\( p\)-value \( \approx 0.003 \)). The cause is the difference of the impression for the tweets between Type 1 and Type 2. Focusing on Figure 3, the largest difference between those is the option (a), the rate of the subject persons who have positive impression for the movie in a tweet of Type 2, is higher than Type 1 by 0.26 (\( p\)-value \( \approx 0.000 \)). Therefore, it can be said that users’ interests are induced by the positive reputation for the movie in a tweet.

Secondly, focusing on the option (2) in Figure 2, the rate of the subject persons who feel that the movie in a tweet of Type 4 would be interesting, is higher than Type 3 by 0.42 (\( p\)-value \( \approx 0.000 \)). Further, focusing on the option (1) in Figure 2, the rate of the subject persons who want to watch the movie in a tweet of Type 4, is higher than Type 3 by 0.09 (\( p\)-value \( \approx 0.004 \)). This is also because the rate of the subject persons who have positive impression for the movie in a tweet of Type 4, is higher than Type 3 by 0.78 (\( p\)-value \( \approx 0.000 \)).

Summarizing the above, it is suggested that the positive reputation for the movie in a tweet has an effect to induce the users’ positive interests on the movie.
B. Verification of Hypothesis 2

As a result, this paper reveals that the newness of a tweet would be one of the factors to induce users’ interests positively. Focusing on the option (c) in Figure 3, some subject persons have an impression that the tweets of Type 4 are on the topic of recent movies, and the tweets of Type 2 are not. Focusing on the option (1) in Figure 2, the rate of the subject persons who want to watch the movie in a tweet of Type 4, is higher than Type 2 by 0.08 (p-value $\leq 0.013$). Further, focusing on the option (2) in Figure 2, the rate of the subject persons who feel that the movie in a tweet of Type 4 would be interesting, is higher than Type 2 by 0.26 (p-value $\leq 0.000$). However, we cannot conclude that the newness of a tweet certainly induces users’ interests positively, because there are other differences of the impression for the tweets between Type 2 and Type 4. As an instance, focusing on the option (a) and the option (e) in Figure 3, some subject persons feel that a tweet of Type 4 introduces the movie more positively, the others feel that a tweet of Type 2 is not on the topic of movies.

Summarizing the above, it would be said that the Hypothesis 2 is partly correct, and that the newness of a tweet would be expected to induce users’ interests positively. However, in this experiment, there is a possibility that these users’ interests are induced by the other factors. As a future work, we need to analyze the relationship with the other factors, and how much the newness of a tweet has an effect to induce interests positively.

We can say the same thing for Type 1 and Type 3. Focusing on the option (b) in Figure 3, some subject persons have an impression that a tweet of Type 3 introduces the movie more negatively than Type 1. Due to this, although some subject persons have an impression that a tweet of Type 3 is on the topic of recent movies, the rate of the subject persons whose positive interests for the movie in a tweet of Type 3 are induced, is almost the same as Type 1.

C. Verification of Hypothesis 3

As a result, it would be said the Hypothesis 3 is partly correct, because when giving 3 posted person’s names, the rate of the subject persons whose interests are changed, is 44%. The remainder of this section discusses what kinds of remarkers give users what kinds of changes of their interests.

Focusing on the option (1) in Figure 4, the remarker who has the most positive influence is Remarker 1. The characteristics of Remarker 1 is shown at the option (e) in Figure 5. Some subject persons have an impression that Remarker 1 has reliability on movies, and the difference between Remarker 1 and the others is statistically significant (p-value < 0.01). Therefore, it is suggested that the remarks of a person who has a wealth of knowledge on the remarked topic, would be one of the factors to induce users’ interests positively.

Focusing on the option (2) in Figure 4, the remarker who does not have any influence is Remarker 3. The characteristics of Remarker 3 is shown at the option (h) and the option (g) in Figure 5. Some subject persons do not know about Remarker 3 and feel that Remarker 3 does not have famousness, and the difference between Remarker 3 and the others is statistically significant (p-value < 0.001).

However, focusing on the option (3) in Figure 4, the option (a), the option (d), and the option (g) in Figure 5, despite the fact that some subject persons know Remarker 2 and feel that Remarker 2 has famousness, the rate of the subject persons whose interests are changed negatively by Remarker 2, is the highest. This is because the rate of the subject persons who like Remarker 2 is lower than otherwise. Summarizing the above, the interests for what the person remarks would change depending on the characteristics of a posted person. And it is suggested that not only famousness and reliability but also “favor” of the posted person are important for changing interest positively.
VII. Conclusion

Our research sets a problem that industries of a genre which a user is not interested in cannot get the user as a new customer, and focuses on the “Influencer,” who is the person that has a large impact on a user’s behaviors. We aim to induce the user’s latent interests of a genre which the user is not interested in, by utilizing the influencer’s remarks that are on the topic of the genre.

This paper made experiments on relationships between “the natural and casual remarks that the remarker does not attempt to advertise items intentionally” and buying emotion of users. We set some hypotheses on the case that the user’s interests on movies is induced by remarks of someone, and reveal the following findings by verifying these hypotheses.

1) Even if it is the natural and casual remarks that the remarker does not attempt to advertise items intentionally, the introduction with the positive reputation for the item and the introduction of the timely item, would induce users’ positive interests for what the remarker says.

2) By various factors such as famousness, reliability and favor for the remarker, users’ interests for what the remarker says would change positively or negatively.

As a conclusion of this paper, the natural and casual remarks that the remarker does not attempt to advertise items intentionally would be effective on inducing users’ interests positively. However, we need to improve the automatic extraction of the remarks that introduce the movie positively, because the rate of the subject persons whose interests of the movie are induced positively by the remarks that are extracted manually, is 76.4%, on the other hand, the case by the remarks that are extracted automatically is 60.4%.

Further, we also need to estimate the influencer for each user in our future works. There is a problem that the influencer for the user does not always remark something which is on the topic of the genre that the user is not interested in. We need to study an influencer’s estimation method which considers influencer’s level (e.g., Friend-Level, Celebrity-Level) and force of influence of each person, depending on users.

References


