Estimating Consumer Inclination for Agricultural Products from Web Browsing History

Riki Tatsuta, Sunao Nakanishi
Graduate School of Information Science and Engineering
Ritsumeikan University, Shiga, Japan
email:{tatsuta, sunao}@de.is.ritsumei.ac.jp

Yusuke Kajiwara, Hiromitsu Shimakawa
College of Information Science and Engineering
Ritsumeikan University, Shiga, Japan
email:{kajiwara, simakawa}@de.is.ritsumei.ac.jp

***Abstract*—In recent years, more and more e-commerce sites sell agricultural products. It should be noted that users have preference of taste. They also have intention on what kinds of agricultural products they get. They consider them at the time of purchase of agricultural products in recommendation on e-commerce sites. There are a number of studies on the automatic extraction of preference of taste that users have from contexts on the web. However, there has been no study that tries automatic extraction of user intentions to purchase agricultural products on e-shopping sites. E-shopping sites without consideration of user intentions are likely to recommend agricultural products far from what users want to buy. Confused users will never visit such e-shopping site, again. We propose a model which estimates intention level of users from web contents of e-commerce sites and web browsing history of users. The proposed method refers to the degree of user interest about a topic calculated from LDA as the topic attention level. Based on the multivariate logistic regression model, the method constructs an intention level model from the topic attention level. An experiment suggests the method can calculate the safety intention level, focusing on price of agricultural products on web pages.**

Keywords–recommendation; inclination; intention; preference; agriculture; e-commerce site; browsing history.

I. INTRODUCTION

More and more e-commerce sites sell agricultural products. Sale of agricultural products on the Internet has several good effects. Farmers can increase their sales. They are also motivated through direct communications with consumers. Supply of high-quality agricultural products provides them with good brand impression [1]. E-commerce brings convenience of direct delivery of agricultural products to consumer houses. The elderly households are too weak to go out shopping food. Households who have small children have no time to go out for shopping. They are eager to buy food with Internet [2]. On the other hand, farmers need efforts to make more consumers visit their own sites to make success in e-commerce selling agricultural products. Their e-commerce sites present information which motivates consumers to purchase their products. In addition to reviews, prices, and quantity like general e-commerce sites, they have information on varieties and palatability of products. They also present cultivation methods to explain their cultivation policy [3].

Information motivating consumers to buy agricultural products varies with preference of consumers. The preference is determined by various factors. For example, preference on cooking with agricultural products ranges from foodstuff, menu, mouthfeel, taste, seasoning and flavor [4]. Apart from these, consumers have intentions on their peculiar ideas such as safety and costs [5]. Some consumers give a higher priority to safety and palatability than to prices, while others mind prices rather than appearance. A combination of preference and intentions of each consumer forms specific inclination which regulates purchase behavior of the consumer. Farmers need marketing activities to target consumers whose inclination meets their cultivation policy. However, Japanese farmers engaging in highly intensive agriculture cannot afford to consider marketing activities because they are busy with cultivating many kinds of crops. One possible solution to support attracting consumers is Search Engine Optimization (SEO), which makes the site appear more frequently in result list of Internet search engines. However, the solutions cannot support the site builders should make consumers interested in the sites so that the consumers can buy agricultural products buy according to their inclination. From the viewpoints of consumers, they want to buy agricultural products from farmers whose cultivation policy meets their own inclination. For example, consumers with safety intentions would read politely safety explanations of products in the e-commerce sites, because they want to purchase safe products as possible, as far as their price permits. They want to get foods free from pesticides. Trading on the current e-commerce never considers coincidence of farmer cultivation policy with inclination of consumers, when customers search farmers optimal to them. In fact, consumers must manually look for agricultural products fitting their needs from a lot of e-commerce sites, struggling with search engines. However, search engines disappoint them many times, introducing sites which handle mismatching agricultural products.

This paper proposes a method to estimate inclination of consumers, in order to recommend appropriate e-commerce sites to them when they search farmers. The method utilizes Web contents on the e-commerce sites and browsing history of customers. The method figures out the distribution of topics on farmer cultivation policies from a specific Web page using the Latent Dirichlet Allocation (LDA). The paper refers the product of the topic distribution and the browsing time of the web pages as the topic attention level. The method constructs an inclination level model using the topic attention level as predictor variables, based on the multivariate logistic regression. An experiment result suggests the method can calculate the safety inclination level.

This paper is organized as follows. Section II discusses the detail of inclination. The related works are also explained in Section II. Section III describes the outline of the proposed considered recommendation system. The proposed method to estimate inclination is also described in Section III. Section IV presents the results of experiments and evaluation. Section V presents the discussion on improvement of the proposed method. Finally, Section VI concludes this papers.

II. Product Information Matching Inclination

Consumers have preference on agricultural products, such as taste, appearance, flavor and so on. The preference is said to differ by age, generation, gender, regions and occupations [6]. In addition, consumers would stick to their peculiar ideas such as safety and costs, when they choose products to buy. Consumers generally determine products to buy with their preference and intentions. The paper refers the combination of preference and intentions of each consumer as inclination. It usually regulates purchase behavior of the consumer.

There are studies on the estimation of consumer inclination [7][8]. For the estimation, the studies oblige consumers to fill up questionnaires and profile in free description. They analyze results with text mining technologies to find user inclination for agricultural products. However, consumer inclination changes depending on fashion and trend. For example, users minding safety, which means products are free from harmful substance to eat, have greatly increased in Japan because of the Fukushima nuclear accident. It is necessary to successively examine consumer inclination because the thinking ways of consumers vary from day to day. It is an unendurable for consumers to fill up questionnaires and profiles many times. A new method is required to extract current inclination of each consumer without bothering the consumer.

III. Recommending Agricultural Products Matching Inclination

We aim at detecting user inclinations from Web browsing log before they decide which farming products they buy. It does not make any sense to analyze huge data obtained after they decide to buy. The log of URLs in Web browsing of users is also insufficient to achieve the aim above. We need to analyze the contents to see what topics in the pages stimulate the users. The LDA on the contents of Web pages they are interested in enables us to identify their topics.

A. Outline of recommendation with extracted inclination

To extract inclination of consumers at the time, we use their browsing history while they look for their favorite agricultural products. We extract consumer inclination from Web pages the consumer visits while he chooses agricultural products in e-commerce sites. E-commerce sites present a variety of Web pages. In the web pages, there are contents which provide stimuli to consumers, because the contents match their inclination. Inclination of each consumer is likely to appear when he selects products. It seems that Web pages vary with consumers because of their inclination. For example, it may be supposed consumers who put importance to safety purchase domestic or organic tomatoes. In this case, they are supposed to investigate description on agricultural chemicals or product places. On the other hand, consumers preferring economy to safety purchases foreign tomatoes, because of low price. In this case, consumers would browse pages of agricultural products after they sort the pages in the decreasing order of the price. It may be suspected that browsing time of Web pages gets long if they have stimulated consumers. In the proposed method, we estimate consumer inclination from topics described in each Web page and the browsing time to read the Web page. In this work, we focus on inclinations: safety, domestic production, locality, economy, and gourmet from the survey results of inclinations on foods [6] and the studies of selection criterion at purchase of agricultural products [9]. As Figure 1 shows, the paper proposes a method to build a model to estimate consumer inclination from browsing history of the consumer to provide him with agricultural product information. At first, the method extracts browsing pages and browsing time from a browsing history when the consumer chooses agricultural products on e-commerce sites. Second, it calculates the degree of how much the consumer has each inclination. It refers to the degree as the inclination level. Finally, the method recommends the agricultural product information which suits his inclination.

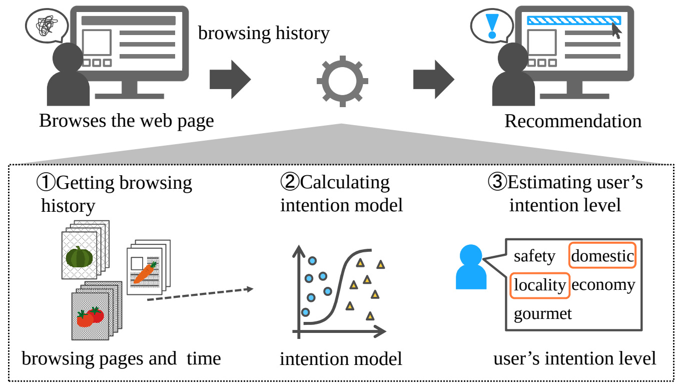

Figure 1. System configuration diagram

There are various agricultural products. For example, organic tomatoes, fruit tomatoes, cherry tomatoes are classified into the category of tomatoes. These categories are referred to as product categories in the paper. Though consumers searching products visit various sites, the method values pages for a specific product category as item pages, rather than pages showing lists of item pages, which are referred to as search pages. Pages irrelevant to e-commerce sites such as Google search and Wikipedia are referred to as other pages in the paper. Item pages are classified into product categories, based on words on their titles and contents. From browsing history of each consumer, item pages are extracted for each product category. As Figure 2 shows, the proposed method builds a model to estimate the inclination of each consumer as follows.

1) The item pages are regarded as a set of words according to the Bag-of-words method [10]. The topic distribution of the item pages is figured out with the LDA. The proposed method calculates the topic attention level with the product of the topic distribution and the browsing time of each item page.
2) The method selects important variables to calculate the inclination level of the consumer.
3) Using the inclination level, the method constructs an inclination model for each consumer.

B. Weighted topic distribution

Through a morphologically analysis, the proposed method divides a set of sentences in item pages into words. It makes

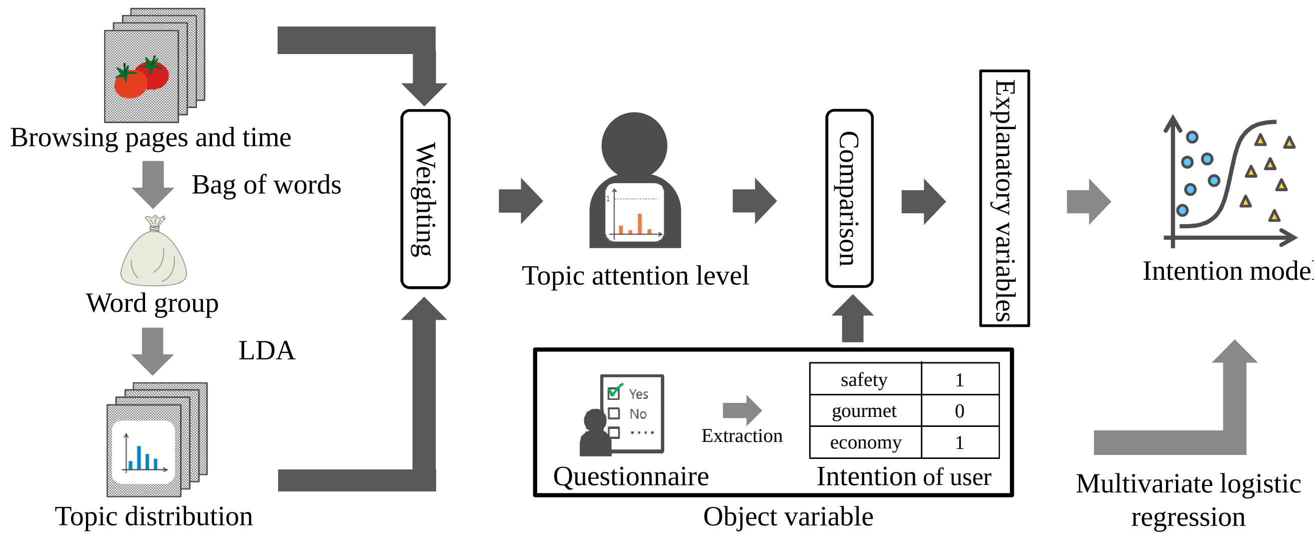

Figure 2. Proposed method

bag-of-words from nouns, adjectives and verbs which seem to be related to inclination. Using the LDA [11], the method acquires combinations of topic distribution and its word group for each page. Though words in the item page are relate to various kinds of inclination, the LDA allows to get topics classified according to each inclination.

The proposed method uses browsing time to determine whether the item page is related to a specific inclination or not. The previous work [12] has clarified a relationship of the browsing time of a Web page to the degree of user interest. LDA is a model which assumes the document is approved from two or more topics. The degree of the interest to a document is that to the topic on the document. The degree of the interest to the topic of the page are related with user's browsing time. The words which relate to several kinds of intention may be included in the e-commerce sites selling agricultural products. However, if each page is weighted by browsing time, user's focus point of words of the page is not considered.

Words which relate to the intention should be classified in one page. This technique subdivides words by a topic distribution of each item page, which enables the acquisition of the degree of each of the user's intentions. Equation (1) shows topic distribution the vectors $\boldsymbol{T_{cp}}$.

$$\boldsymbol{T_{cp}} = [T_{cp1} \quad \cdots \quad T_{cpk} \quad \cdots \quad T_{cpK}]^{\mathrm{T}} \tag{1}$$
$$(1 \leq c \leq C), (1 \leq k \leq K), (1 \leq p \leq P_c)$$

where k, c, and p represent a specific topic, a specific product category, and a specific page, respectively. K and C are the number of topics and product categories, respectively. P_c is the number of item pages which are contained in the c-th product category. T_cpk is the k-th component of topic distribution of the p-th page of the c-th product category. Equation (2) shows vector I_c. Each component of it is a sum of T_cp weighted by the browsing time of each page in each product category.

$$\boldsymbol{I_c} = [I_{c1} \quad \cdots \quad I_{ck} \quad \cdots \quad I_{cK}]^{\mathrm{T}} = \sum_{p=1}^{P_c} (\boldsymbol{T_{cp}}) d_{cp} \tag{2}$$

I_{ck} is the weighted vector component relating to the k-th topic of the c-th product category. d_{cp} is the browsing time of the p-th page belonging to the c-th product category. The more pages belonging to the same commodity category the consumer browses, the more I_{ck}grows. It is necessary to normalize I_{ck}, because the browsing time and the browsing count vary with consumers. Equation (3) shows $S(I_{ck})$ which is calculated with normalized I_{ck}.

$$S(I_{ck}) = \frac{I_{ck}}{|I_c|} \tag{3}$$

In this work, we refer to $S(I_{ck})$, which is extracted from each product category as the topic attention level.

C. Selection of important variables

In Section III-C, we account for the explanatory variables of the making model as shown in Section III-D. The topic with difference of topic attention by the presence of the intention is extracted. Equation (4) shows $G\{n_{ck}\}$, the set of the topic attention level of each consumer.

$$G\{n_{ck}\} := \{S(I_{ck})_1, \cdots, S(I_{ck})_u, \cdots, S(I_{ck})_U\} \quad (1 \leq u \leq U) \tag{4}$$

Whole users are divided into 2 groups: one with a specific inclination and the other without it. In Equation (5) and (6), $X\{n_{ck}\}$ and $Y\{n_{ck}\}$ show consumers with the n-th inclination and without it, respectively.

$$X\{n_{ck}\} := \{G\{n_{ck}\}|P(\{G_{ck}\})\} \tag{5}$$
$$Y\{n_{ck}\} := \{G\{n_{ck}\}|Q(\{G_{ck}\})\} \tag{6}$$

P is the condition that the consumer does not has the inclination. Q is the condition that the consumer has the intention. n and N corresponds to the n-th inclination and the number of it. n_{ck} is the k-th topic of the c-th product category to correspond to the n-th inclination. The method aims to build a model to estimate inclination of unknown consumers. It assumes inclination of several consumers is acquired beforehand by questionnaires or interviews. Though it obliges some efforts on some consumers, there is no burden on unknown consumers. The method calculates difference of the topic attention level of the group with the inclination and that without the inclination. Let us focus on the difference of the topic attention levels of the two groups. A large difference in a topic indicates the group with the inclination pays their attention on the topic, while the other group pays no attention on it. To examine the difference of the topic attention level, the method uses median values resistant to noises. Equation (7) and (8) calculate the median values of the group with the inclination and the group without it.

$$MedianX(n_{ck}) = \begin{cases} x_{(a+1)/2}, & if\ a\ is\ odd. \\ \frac{1}{2}(x_{a/2} + x_{a/2+1}, & if\ a\ is\ even. \end{cases} \quad (7)$$

$$MedianY(n_{ck}) = \begin{cases} y_{(a+1)/2}, & if\ a\ is\ odd. \\ \frac{1}{2}(y_{a/2} + y_{a/2+1}, & if\ a\ is\ even. \end{cases} \quad (8)$$

Equation (9) shows $diff(n_{ck})$ which is the difference of $MedianX(n_{ck})$ from $MedianY(n_{ck})$.

$$diff(n_{ck}) = MedianX(n_{ck}) - MedianY(n_{ck}) \quad (9)$$

Equation (10) shows $Diff_n$ which is the set of the difference of the median for the n-th inclination.

$$Diff_n := \{diff(n_{11}), \cdots, diff(n_{CK})\} \quad (10)$$

It seems that the median with great difference is an important value to characterize a group of users who do not have intention and a group of users who have intention. This method picks out top R of large median difference in a positive direction as $High_n$ and lowest R of large median difference in a negative direction as Low_n in $Diff_n$. Topics corresponding to $High_n$ is focused by the group with the n-th inclination, but neglected by the group without it. Topics corresponding to Low_n mean the opposite case.

D. Construction of inclination model

The proposed method builds a model to estimate the consumer intention level. The explanatory variable necessary for the model is a topic attention level which belongs to $High_n$ and Low_n in III-C. The object variable is the inclination of each consumer. The method refers to it as an inclination level model. The method applies the multivariate logistic regression model, to make a good inclination level model. To improve the performance of the model, the method excludes explanatory variables belonging to $High_n$ and Low_n in a stepwise way.

The intention model distinguishes presence of a specific inclination of an unknown consumer from his topic attention level. The topics on each Web page are obtainable beforehand, applying the LDA to the contents on e-commerce sites in advance. What is required for the on-line calculation for the inclination level of the consumer is only his browsing time.

IV. Experiment

We experimented to verify the utility of the proposed method. The experiment aims at

- validation of efficiency of difference that user browses topic by presence of the intention of the user, and
- verification of validity of intention level model.

A. Experiment method

We experimented about the purchase of agricultural products using e-commerce sites. The research participants were 17. We acquired the browsing history of the research participant from beginning to the end of the experiment. We had to make the research participants unaware of the costs because that would have influenced their selection. We laid out a situation as follows “I want to celebrate grandmother’s sixty-first birthday with about 10 relatives. My mother said that I should cook for the grandmother. You don’t need to think about money, because I’ll pay. I thought that I will try to cook in grandmother’s house, but there is no supermarket nearby. So, I decided to prepare foodstuff using e-commerce sites.”. In addition, the grandmother said “I want to eat home-made dish because I dislike ready-made one. I want you to buy rice you recommend.”. Below are the proposed dishes: salad, rice cooked with matsutake mashroom, boiled dishes, and meat dish. The material necessary for dish is shown as follows: tomato, cucumber, matutake, rice, radish, taro, chicken, and meat. We presented the amount of the foodstuff necessary for the dish and the recipe of the dish to research participants. We specified e-commerce sites to buy agricultural products. We permitted them to use retrieval engines on Internet to gather information. Each of the user’s intentions was acquired in the questionnaire. We made a questionnaire, creating an association between the information that the research participants are interested in e-commerce site and each intention referring to result investigation of intention [6]. It seems that the user who has gourmet intention takes an interest in palatability. For example, information is presented on e-commerce sites such as ”sugar concentrations are higher and sweeter than another”, ”taste becomes better than another by the cultivation method of original” and ”agricultural products harvesting in the morning are very fresh”. The user who has a gourmet intention is attracted by the deliciousness or the sugar concentrations. It is seems that the user who has an economy intention is attracted by low price or large quantity. For example, information about price and quantity are presented on e-commerce sites. Therefore, the user who has an economy intention will browse a page that contains a price or quantity. However, we only asked research participants about whether or not they are attracted by price because of the quantity of agricultural products that they should buy were specified. Questionnaire items for the safety intention, the domestic intention and the locality intention are clear. After buying agricultural products, we send out questionnaires of Table I. in three grades (yes, no, unknown).

TABLE I. QUESTIONNAIRE ITEM

Intention	*Questionnaire item*
Gourmet intention	Q1. Did you interest in deliciousness of the testis? Q2. Did you interest in freshness?
Safety intention	Q3. Did you interest in level of safety?
Domestic intention	Q4. Did you interest in domestic production?
Locality intention	Q5. Did you interest in producing area?
Economy intention	Q6. Did you interest in value?

B. Evaluation

1) Discriminating existence or nonexistence of user’s intention by an intention level: Data to acquire topic attention level are documents of item pages which are browsed by

17 research participants. The number of commodity page is 503. We classified those pages in the commodity category. In addition, because the result that research participants bought food stuff of meat dish are divided into pork and beef, we classified them as a same commodity category. Sometimes, information not related to the retrieval intention of the user has been presented in item page. For example, there are information on recommended commodity and the commodity ranking of peculiar to sites. The documents were excluded since these documents are not relate to user's intention. The review which contain on the item page was also excluded.

2) Extraction of topic from the item page: We applied the method explained in III-B to classified item pages. Among words which appear frequently, the intention is not considered from words and symbol of the commodity category which does not relate to the intention. A bag-of-words was prepared, excluding these words. LDA is applied to the bag-of-words using collapsed Gibbs sampling [13]. The scalar value of the Dirichlet hyperparameter for topic proportions is 0.1. The beta hyperparameter for each entry of the block relations matrix is 0.1. The number of trials is 30000. Among the 5 patterns (2, 4, 6, 8, 16 topics), 4 topics presented the best performance. The top 5 words were selected from each of topics.

3) Acquisition of the user's intention: Research participants were responded to a questionnaire of Table I. If user answered YES to the questionnaire of palatability and freshness, the user is regarded to have gourmet intention.

C. Result

The method explained in paragraph III-B was applied to extract the topic. Words were labeled based on generality.

1) Extraction of topic attention level: The method explained in paragraph III-C was applied to extract the topic attention level. Figure 3 shows the median of extracted topic attention level. Table II shows the result of the questionnaire about an intention. ○ is yes. X is no. △ is unknown.

TABLE II. RESULT OF THE QUESTIONNAIRE ABOUT AN INTENTION

User	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*
Q1.	○	○	○	○	○	○	○	X	○	○	○	○	△	○	△	○	X
Q2.	○	○	X	○	○	○	○	△	X	X	△	X	○	○	○	○	△
Q3.	△	X	○	○	○	○	△	△	○	X	X	X	X	○	○	○	○
Q4.	○	○	○	○	△	○	○	○	○	X	○	○	○	○	○	○	○
Q5.	○	○	X	○	○	○	○	△	○	X	X	○	X	X	○	○	○
Q6.	○	X	○	X	○	○	○	X	△	○	○	X	○	○	X	X	○

2) Discriminating existence or nonexistence of user's intention by an intention level: We have verified whether we can discriminate user's each intention existence from user's intention level by the intention model which is explained in paragraph III-D. The Table III shows result of the multivariate logistic regression model. The significance level is less than 5%. The significance level (0.042) of a topic intention relating a value about a safety intention is extracted from result of the multivariate logistic regression model. The significance level (0.042) is a significant value. This result suggests that a safety intention is connected with topics of price. The significance level of each intention except a safety intention has been low. The significant value to estimate user's intention was not found by an intention model.

TABLE III. RESULT OF THE MULTIVARIATE LOGISTIC REGRESSION MODEL

Intention	*Label*	*Estimate*	*Std.Error*	*z value*	$Pr(>\lvert z\rvert)$
domestic	agricultural chemical cultivation method	44.99	20068.94	0.002	0.998
	brand	-494.47	168178.07	-0.003	0.998
	domestic, high quality cook	1074.23	358333.88	0.003	0.998
	deliciousness	-898.71	305721.58	-0.003	0.998
	AIC	12.773			
gourmet	cook	30.8814	21.4463	1.44	0.1499
	AIC	18.025			
economy	agricultural chemical cultivation method	-7001	741189	-0.009	0.992
	deliciousness	-23338	2467722	-0.009	0.992
	locality, value	5922	626767	0.009	0.992
	AIC	8			
safety	domestic	2.802	1.984	1.412	0.1579
	value	6.75	3.32	2.033	0.042
	AIC	20.537			
locality	locality	-3.823	2.396	-1.596	0.111
	deliciousness	8.476	6.372	1.33	0.183
	AIC	20.306			

V. DISCUSSION

A. Discussion of the intention model

The significant value to estimate user's intention has not found by an intention model. In this work, it is assumed user's intention is relate to browsing pages of time and contents. The longer users browse pages, the more they are interest in the pages. However, the time changes, depending on how to browse the page. The weighting is considered as one of the causes which makes the significance probability low. The proposed method decides weight without considering how to scan the page. It is investigated that the inspection time changes depending on how to scan the page. The examination method focuses on how to browse page using tab or click. The Table IV shows the counts of how to browse page using tab or click. If the frequency in which the page is opened with a new tab is smaller than that with the current tab, we call those users a "Click type", otherwise "Tab type". The click type are 2, 3, 4, 8, 10, 13 and 14, while the Tab type is 5, 6, 7, 9, 11, 12, 15, 16 and 17. The Table V shows the difference between the average of the browsing time from the average of the browsing frequency. It can be said that browsing time and counts of the tab type are larger than that of the click type. The difference at the browsing time appears in how to scan the page. It is necessary to consider the method of the scanning of the individual when weighting to topics using browsing time. It can be expected that the higher degree of fit model is constructed, considering the method of the scanning.

TABLE IV. THE COUNTS OF HOW TO BROWSE PAGE USING TAB OR CLICK

User	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*
Click	59	1	0	1	42	18	13	1	18	0	13	21	0	1	16	20	18
Tab	9	19	21	66	1	0	3	39	0	24	0	1	53	36	1	2	0

TABLE V. THE AVERAGE OF THE BROWSING TIME AND THE AVERAGE OF THE BROWSING FREQUENCY

	Tab type	*Click type*
Average of browsing time(second)	51.945	42.354
Average of browsing frequency(second)	1.935	1.457

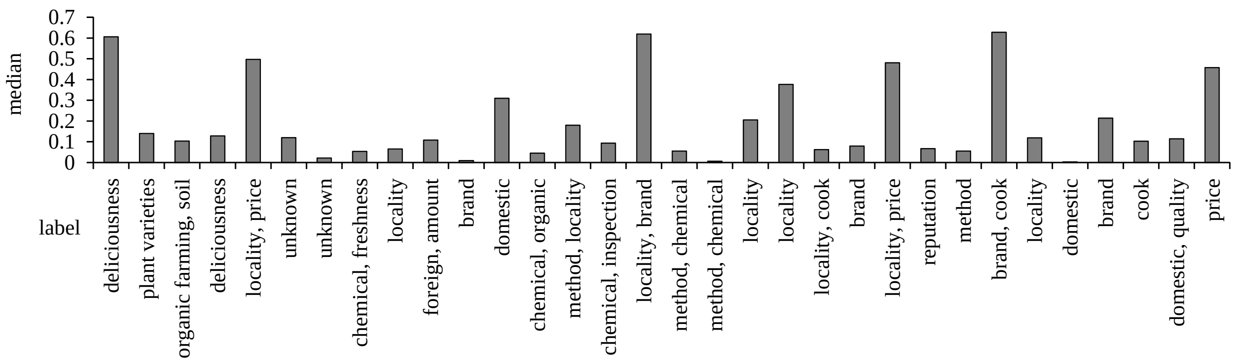

Figure 3. Median of extracted topic attention level

B. Acquisition of Web browsing history

The proposed method needs access to the web history of users. It might sound a rather demanding assumption. Users agreeing to the method are assumed to install an application to collect Web browsing histories. It is a trade-off between their advantages and disadvantages. We should consider two kinds of disadvantages: a burden to install the application, and a risk to expose their privacy. On the other hand, the users are released from stresses coming from recommendation of irrelevant information. We thought the advantages are greater than the disadvantages.

VI. CONCLUSION

The present paper estimated an intention level to present the information based on consumer intention on e-commerce sites. In this method, we extracted item pages which are contained each commodity using browsing history. We prepared a bag-of-words from documents which are contained item pages. We applied LDA to the bag-of-words to extract topic distribution. We extracted topic attention level by weighting browsing time and topic distribution. We constructed the intention model using the multivariate logistic regression model by an important topic attention level and user's intention. The user's intention is estimated from a topic attention level with the intention model in an experiment. The significance level is less than 5%. The significance level (0.042) of a topic intention relating a value about a safety intention was extracted from result of the multivariate logistic regression model. To improve fit degree, we consider transition of the pages. In addition, the questionnaire items do not necessarily correspond to separate categorize item pages. We plan to review the questionnaire.

REFERENCES

[1] Japan Finance Corporation: Survey of sixth sector industrialization in agriculture, https://www.jfc.go.jp/n/findings/pdf/topics_111202_1.pdf [retrieved: March, 2015] (in Japanese).

[2] JC general research: Survey results of the consumption behavior of vegetables and fruits, http://www.jc-so-ken.or.jp/pdf/agri/investigate/141203_01.pdf [retrieved: March, 2015] (in Japanese).

[3] Agriculture, forestry and fisheries. Information Exchange Network Business: Survey results of awareness and intention on the production and distribution information of vegetables, http://www.maff.go.jp/j/finding/mind/pdf/20040820cyosa.pdf [retrieved: March, 2015] (in Japanese).

[4] Ministry of Agriculture, Forestry and Fisheries: Japanese preference and taste, http://www.maff.go.jp/j/keikaku/syokubunka/culture/mikaku.html [retrieved: March, 2015] (in Japanese).

[5] Japan Finance Corporation: Survey of food intention: https://www.jfc.go.jp/n/findings/pdf/topics_140911a.pdf [retrieved: March, 2015] (in Japanese).

[6] Yamaguchi. K, and Takahashi. C, Studies on the Preference of Foods : Part 2 : Relationship between the attribute and the preference of food, Cookery Science 15.2, 1982, pp. 104-113.

[7] Isojima. A, Analysis of Consumers Needs for Agricultural Products by Using Text Mining, National Agriculture and Food Research Organization Vol.16, No.4, 2010, pp. 38-42.

[8] Isojima. A, Analysis of a Consumer Questionnaire Pertaining to Rice by using Text Mining. Agricultural Information Research. Vol.15, No.1, 2006, pp. 49.

[9] Isojima. A, Consumer Perceptions and Selection Standards Rice Purchases : An Investigation into Consumer Considerations, Japanese Journal of Farm Management Vol. 36, No. 3, 1999, pp. 22-32.

[10] C. D. Manning, H. Schutze, "Foundations of statistical natural language processing." MIT Press, Cambridge, MA, 1999, pp. 235-260.

[11] D. M. Blei, A. Y. Ng, M. I. Jordan, "Latent dirichlet allocation." the Journal of machine Learning research 3, 2003, pp. 993-1022.

[12] Morita. M, and Yoichi. S, "Information filtering based on user behavior analysis and best match text retrieval." Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval. Springer-Verlag New York, Inc., 1994, pp. 272-281.

[13] T. L. Griffiths, M. Steyvers, "Finding scientific topics." Proceedings of the National academy of Sciences of the United States of America 101.Suppl 1, 2004, pp. 5228-5235.