Open Access Journal of Information Systems (OAJIS)

Available at http://is.its.ac.id/pubs/oajis/index.php

The Third Information Systems International Conference

A fuzzy ad selector model based on browsing history

Titah Yudhistira[a,b*], Ratna Mutiara Dewi[a]

[a]*Industrial Engineering Department, Institut Teknologi Bandung, Ganesha 10 Bandung 40132, Indonesia*
[b]*Engineering Management Department, Institut Teknologi Bandung, Ganesha 10 Bandung 40132, Indonesia*

Abstract

Nowadays, online advertising is regarded as the most prospective form of advertising. One kind of online advertising is banner advertising. However, click through rate (CTR), which measures banner effectiveness, has been declining as low as 0.1 percent. CTR is important for publishers since the payment from advertisers is based on the number of clicks. Previous studies show that showing internet users with banners that match their online behavior will increase CTR. This behavioral targeting can be implemented by classifying the users based on their click-stream data from their navigation and present the banners to users whose history behavior indicates high interest on advertised products. This paper proposes a web ad selector model in order to determine suitable banner advertising to the audience using their browsing history (web server log file). This web ad selector consists of two models: fuzzy inference model and web ad matching model. From the face validity our web ad selector model always includes the banner ad selected by the expert, with matching score average of 75 percent. This proves that our proposed model has a good performance.

Keywords: Type your keywords here, separated by semicolons ;

1. Introduction

With the increase of internet access, the Internet has attracted more advertisers with its online advertising. One form of online advertising is banner advertising. Despite its simple form, banner advertising accounts for a major share in internet advertising spending. In 2010, US banner advertising generated $ 1.9 billion and accounted for 26% of total internet advertising. The amount spent increased by 24% from the previous year [1]. One commonly used measure to evaluate banner effectiveness is click-through rate (CTR). CTR is the average number of clicks by the audience of a banner that brings them to

* Corresponding author. Tel.: +62-22-2510680; fax: +62-22-2510680
E-mail address: titah@mail.ti.itb.ac.id.

the target website [2]. However, CTR has been declining in recent years and recorded as low as merely 0.09 percent [3]. CTR is important for publishers, i.e. web sites where the advertising are placed, since the payment from advertisers is based on the number of clicks. The higher the number of clicks, the higher the publisher revenue.

Dreze and Hussherr [4] found that low CTR was caused by the fact that most internet users avoided seeing the banners because they were not interested in the banners. Based on this finding, one solution to increase CTR is by using targeting. Targeting aims at presenting banners selectively to individuals who most probably will give positive response. However, the biggest challenge is to identify the individuals.

There are several approach in targeting the audience, i.e. data-driven demographics, behavioral, channel, geographic, and re-targeting [5]. When the information is restricted to browsing history, then behavioral targeting is the only feasible option. Behavioral targeting is based on audience past online behavior reflected in their browsing history. Past online behavior can show the audience specific interests.

In order to a banner can be directed to the right audience, the behavioral targeting tool should be able to identify the audience, to predict and to understand the audience preference correctly so that it can deliver the right banner to the audience during their online session. In this paper, we propose a web ad selector model that can personalize banner content to the audience based on their interest and preference.

In short, to achieve the goal, our web ad selector model uses the browsing history using recorded cookies in audience's browser device. A fuzzy inference model based on fuzzy rule based system (FRBS) Mamdani is employed to produce audience interest score for each advertising category. FRBS Mamdani is a simple and powerful tool that are commonly applied to solve classification problem [6]. Then, this score together with the fuzzy characteristic score in an advertising category for a brand is then processed using a web ad matching using Hamming distance to produce a ranking of banners to be presented to a particular audience.

In the subsequent sections, we explains the model starting with preprocessing data stage, fuzzy web ad selector model, web ad matching model, and concluded with model implementation and analysis.

2. Preprocessing or input data stage

Each browsing history data collected by the cookies have information on unique ID, URL reference, and time of access. These data are then sorted by unique ID to merge data with the same unique ID so that the access pattern of each unique ID can be revealed. Next, the URL reference is classified into 27 categories of websites, i.e. culture and entertainment, automotive, education, family, etc. Since there is no specific rule for this categorization, we use the categorization of the company in our case study. The output of this stage is the frequency and proportion of each website category visit by each unique ID.

The next stage is the fuzzification stage where the frequency of access is translated into degree of membership on fuzzy sets in the form of linguistic variables [7,8] to represent the pattern of each unique ID website visits on each website category. For simplicity, we employ triangular fuzzy number to represent the visit patterns, i.e low, medium, and high (Fig. 1). The membership function of access frequency proportion (x) on the fuzzy linguistic variables is can be defined as a uniform partition using this set of equations:

$$\mu_{Low}(x) = \begin{cases} 1 & if\ x \in [0,0] \\ \dfrac{x-0,375}{0-0,375} & if\ x \in [0;0,375] \\ 0 & otherwise \end{cases} \tag{1}$$

$$\mu_{Medium}(x) = \begin{cases} \dfrac{x-0,125}{0,5-0,125} & if\ x \in [0,125;0,5] \\ \dfrac{x-0,825}{0,5-0,825} & if\ x \in [0,5;0,825] \\ 0 & otherwise \end{cases} \tag{2}$$

$$\mu_{High}(x) = \begin{cases} \dfrac{x - 0{,}625}{1 - 0{,}625} & if\ x \in [0{,}625; 1] \\ 1 & if\ x \in [1; 1] \\ 0 & otherwise \end{cases} \tag{3}$$

Another fuzzification process in this stage is regarding the advertised brands. Advertising experts classified all available banners into 17 categories of brands, i.e. banking and finance, e-channel, automotive, household appliances, equipments, electronics, etc [9]. Then, fuzzification is employed to determine the membership degree of how each brand belongs to each brand category fuzzy set.

3. Fuzzy inference model

The fuzzy inference model consists of three sub-model, (i) knowledge base model, (ii) fuzzy rule based model, and (iii) defuzzification model.

3.1. Knowledge base model

The knowledge base model is built by defining fuzzy rules by advertising experts. These fuzzy rules need input of whether there is relationship between each website category and each brand category. The relationship is determined by advertising expert, and an excerpt of the relationship is preseented in Table 1 as an example. If there is a relationship between a website category and a brand category, then the relationship is translated into three fuzzy rules. Each fuzzy rule contains the same value of the website category linguistic variable (as the antecedent variable) in the *IF* part and brand category linguistic variable (as the consequent variable) in the *THEN* part. In other word, if the value of the antecedent linguistic variable is low, then the value of the consequent linguistic variable is also low.

As an example, from the relationship presented in Table 1, we can define nine fuzzy rules as follows:

1. IF culture and entertainment is low THEN electronics is low
2. IF culture and entertainment is medium THEN electronics is medium
3. IF culture and entertainment is high THEN electronics is high
4. IF automotive is low THEN banking and finance is low
5. IF automotive is medium THEN banking and finance is medium
6. IF automotive is high THEN banking and finance is high
7. IF automotive is low THEN automotive is low
8. IF automotive is medium THEN automotive is medium
9. IF automotive is high THEN automotive is high

3.2. Fuzzy rule based system

We adopt fuzzy rule based system (FRBS) Mamdani [10] to obtain interest score on each brand category for each unique ID. For this FRBS Mamdani, the input variables (antecedent) are all website categories and the output variables (consequent) are all brand categories. We then define the mathematics operations for AND, OR, implication and aggregation fuzzy rules. The fuzzy rule mathematics operation for AND relation is MIN operator, for OR relation is MAX operator, for implication is MIN operator, and for aggregation is MAX operator. For the execution of the FRBS Mamdani we employ *Fuzzy Logic Toolbox* within *Matlab* 7.12.0 (R2011a).

Table 1. Example of relationships between website categories and brand categories based on expert judgement

Website category	Brand category				
	Banking & Fin.	e-channel	Automotive	Home app.	Electronics
Culture and entertainment					√
Automotive	√		√		

3.3. Defuzzification model

The final output of the fuzzy web ad selector model is each unique ID's interest score on each brand category. We use center of gravity (COG) method as the defuzzification method to translate the fuzzy consequence variables into interest scores. COG is used considering its computational efficiency [11].

3.4. How the system works: an example

Below, we explain how the fuzzy inference system works.

1. for each unique ID, input the access frequency (proportion) for each website category into FIS
2. FIS will calculate the degree of membership from the input for each website category, to be used as antecedent variable in the fuzzification stage
3. FIS will determine the fitness value for each fuzzy rule using MIN operation (∧) below:
$$W_k = \mu_{A_1}(x_1) \wedge \mu_{A_2}(x_2) \wedge \ldots \wedge \mu_{A_i}(x_i) \tag{4}$$
where:
W_k : fitness value for k-th fuzzy rule
x_i : frequency of access
$\mu_{A_i}(x_i)$: membership degree of i-th antecedent variable (website category)
4. FIS will imply the fitness value onto consequence from the rules by using clipping process. Figure 2 illustrates the clipping process for consequent variable Electronics on all (nine) fuzzy rules defined previously.
5. FIS will aggregate the output from the clipping process of the consequent variable on all rules by using MAX (∨) operation below:
$$\mu_C(z) = \bigvee_{j=1}^{rule\ \#} \left[W_j \wedge \mu_{C_j}(z)\right] \tag{5}$$
where $\mu_C(z)$ is consequent z (brand category z) fuzzy set. Because in this example all clipping process for each of the nine fuzzy rules are the same, then the FIS result is exactly the same as Fig. 2.

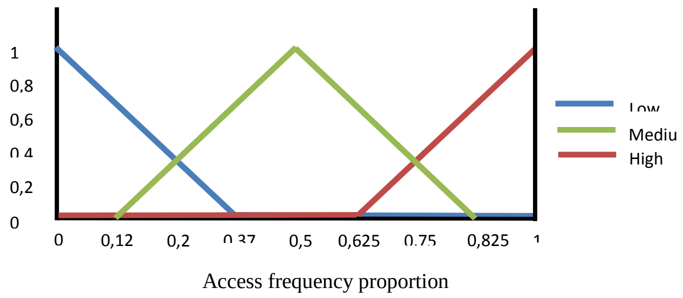

Fig. 1. A triangular fuzzy number to represent visit pattern

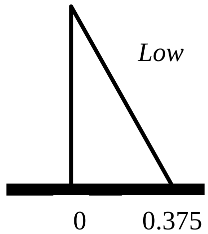

Fig. 2. An illustration of the clipping process for consequent variable Electronics on the nine fuzzy rules

6. Finally, FIS will calculate interest score for each consequent variable and each unique ID through center of gravity defuzzification:

$$\mathrm{COG}(i) = \frac{\sum_x \mu_i(x) \times x}{\sum_x \mu_i(x)} \qquad (6)$$

where

$\mathrm{COG}_z\,(i)$: brand z interest score (crisp) for i-th unique ID

$\mu_C(z)$: consequent *z* (brand category *z*) linguistic fuzzy set

z : interest score

4. Web ad matching model

In the web ad matching model, banners advertising some brands will be matched with each unique ID. The banner advertising for each brand has a fuzzy characteristic in the sense that a single brand can be related, to some degree, with different categories of brands. For example, Samsung Galaxy Note can be categorized into electronics, telecommunication, retail, and e-channel brand category fuzzy set with different degree of memberships.

Basically, this model determines the similarity between the unique IDs (and their interest score on each brand categories) and the banners (and their fuzzy characteristics on brand categories) by calculating the Hamming distance using the formula:

$$d(A, B) = \sum_{i=1}^{n} |\mu_A(x_i) - \mu_B(x_i)| \qquad x_i \in X \qquad (7)$$

where:

d(A,B) : Hamming distance

$\mu_A(x_i)$: unique ID *i's* interest score on website category A

$\mu_B(x_i)$: fuzzy characteristic score of brand B on website category A

The obtained Hamming distances are then sorted from the smallest to the greatest for each unique ID. The smaller the distance, the more fit the banner to be presented to the unique ID. In other word, a banner advertising of the brand with the smallest Hamming distrance will be prioritized to be displayed to the unique ID.

5. Model implementation, analysis, and future extension

In this research, model implementation and testing were done off-line. Data were obtained from an online advertising network company in Jakarta. The data sample are browsing history from 150 unique IDs during the period of October and November 2013. We validated the web ad selection for a particular unique ID resulted from our web ad selector model using the selection made by advertising experts. This face validity check using experts were done by asking advertising experts to determine what banner advertisings that were suitable for a particular audience based on his/her web access pattern. We provided the experts with browsing history from ten unique IDs (chosen randomly) and seven brands to be matched. The seven brands represented brands that can be categorized into different brand categories.

From the face validity our web ad selector model always includes the banner ad selected by the expert, with matching score average of 75 percent. Although only in two occassions our model has the top rank that is the same as the expert's top rank, considering most people will visit more than one websites in one online session, our model will almost always present the suitable banner advertisings, i.e. the banner advertisings that suit the interest of the audience.

In conclusion, in this paper we have presented a fuzzy web ad selector model that can be used by web publishers or online ad network companies to determine the suitable banner advertising to be displayed to their audience using their browsing history. Once the fuzzy inference model (for inferencing unique IDs website interest scores) and web ad matching model (to match brand advertising characteristics with unique ID interest) are built, the model is simple because it uses only audience browsing history. The model implementation and validity test show that the model works reasonably well, although more implementation is needed to ensure that the model does not overfit to current cases and data.

For future research extension, we are currently developing another approach using customer segmentation methods. Here, instead of determining match advertising individually, the method will determine it based on group of audience that show similar interest (based on their clicking frequencies on each website category). In this way, it is expected that modeloverfitting, if there is any, can be avoided.

References

[1] Interactive Advertising Bureau.. IAB Internet Advertising Revenue Report 2010 Full Year Report; 2010. Available at http://www.iab.net/media/file/IAB_Full_year_2010_0413_Final.pdf

[2] Novak, T., and Hoffman, D. L. New metrics for new media: Toward the development of Web measurement standards. *World Wide Web Journal* 1997, ***2***(1), 213-246.

[3] Double Click. 2009 Year-in-review Benchmarks: A DoubleClick Report; 2010 Available at http://static.googleusercontent.com/external_content/untrusted_dlcp/www.google.com/en//doubleclick/pdfs/DoubleClick-07-2010-DoubleClick-Benchmarks-Report-2009-Year-in-Review-US.pdf

[4] Dreze, X. and Hussherr, F.X. Internet Advertising: Is Anybody Watching? *J. of Interactive Marketing* 2003; ***17***(4), 8-23.

[5] Jandal, H. *Display Advertising : The Billboards of the Web,* WSI White Paper; 2011.

[6] Fernández, A. and F. Herrera, Linguistic fuzzy rules in data mining: follow-up Mamdani fuzzy modeling principle, in: E. Trillas, P.P. Bonissone, L. Magdalena, J. Kacprycz, (Eds.), *Combining Experimentation and Theory. A Hommage to Abe Mamdani*, Berlin: Springer, 2012.

[7] Zadeh, L.A. The concept of a linguistic variable and its application to approximate reasoning. *Information Sciences* 1975; **8**, 199-249

[8] Mandal, S.N., Choudhury, J.P. and Chaudhuri, S.R. In search of suitable fuzzy membership function in prediction of time series data. *Int. J. of Computer Science Issues* 2012; **9**(3), 293-302.

[9] Top Brand Award, 2014. Available at http://www.topbrand-award.com/top-brand-survey/survey-result/top_brand_index_2014.

[10] Iancu, I. A Mamdani Type Fuzzy Logic Controller. In Dadios, E. P. (ed.), *Fuzzy Logic – Controls, Concepts, Theories and Application*, Vienna: InTech Publishers, p.325-350, 2012

[11] Klir , G. and Yuan, B. Fuzzy Sets and Fuzzy Logic. Prentice Hall, Englewood Cliffs, 1995.