

Fuzzy logic in automatic analysis of Google Analytics data

Sergiu Chilat

Abstract

This article presents a method for automation of the process of data analysis generated from Google Analytics, using fuzzy logic. After the method implementation, the user will get the most useful and suitable content depending on the age he/she has. Since Google Analytics doesn't provide any automatic analysis tools, implementing this model would bring an appreciable number of advantages, like time saving.

Keywords: Fuzzy logic, web analytics, automatization, self-tuning, fuzzyfication.

1 Introduction

One of the most popular search engines – Google, provides specialized tools – Google Analytics(GA). It allows the storage of the data about the website users, such as: geographic location, age, preferences and interests, etc. [1]. This information is used to optimize the website, but at the moment there are no available methods or tools for automating the decision-making process, so this responsibility is taken by the SEO (Search Engine Optimization) administrator. There are frequent instances when the SEO performs the same routine of operations repeatedly, leading to a decrease in yield and productiveness and also to the adoption of poor quality decisions. In this article, the possibility of implementing fuzzy logic to automate the decision making process in order to save time and free the administrator of such routine activities, will be investigated. One of the GA compartments is the users

age, which will be used in this research and based on it, a method of automatic data analysis and optimal decision making will be proposed, using the maximization function in the defuzzification process.

2 The fuzzyfication process

In the section bellow, Google Analytics gathered the collected data into 6 groups, G_1 – G_6 (input variables):

- G_1 – from 18 to 24 years old;
- G_2 – from 25 to 34 years old;
- G_3 – from 35 to 44 years old;
- G_4 – from 45 to 54 years old;
- G_5 – from 55 to 64 years old;
- G_6 – above 65 years old.

But, the information on the site is divided into 3 groups [2]:

- FG_Y – for young people;
- FG_M – for middle-age people;
- FG_O – for old people.

Since there are no clear criterias of division between these age categories, the dividing into 6 groups (G_1 – G_6) which Google Analytics offers, can't be used directly, being necessary to implement an algorithm that could determine to which group of those 3 (FG_Y , FG_M , FG_O) does the user belong (Table 1).

Table 1. The representation of fuzzy age distribution

	18	24	34	44	54	64	65+
Young	1	0.75	0.25	0	0	0	0
Midle	0	0.25	1	1	0.25	0	0
Old	0	0	0	0.25	0.75	1	1

Here we can apply the Fuzzy logic [3] to present a specific content depending on the group of users. This will define the linguistic variable

age, that can receive one of the following values: young, middle-aged, elderly (Table 1). Basing on the data from the Table 1, the diagram of fuzzy age distribution is proposed (Fig. 1).

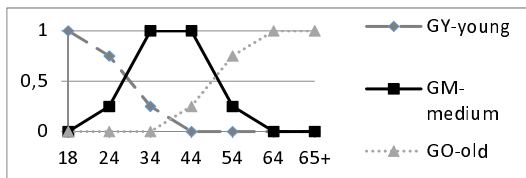


Figure 1. Fuzzy age distribution diagram.

3 Rule base definition

As it can be seen from Fig. 1, a user is “very” young if he is 18 years old (maximum value 1), “still” young at the age of 24 (value 0.8), “quite” young (value 0.2) at age 34 years and at the age of 44 years not young anymore (value 0). The same approach will be applied to the other two groups: middle-aged and elderly. Hereby, a rule can be created [3], which will transform the input data into output data as follows:

IF age = young THEN content = for young

IF age = medium THEN content = for medium

IF age = old THEN content = for old

In this way, the system receives at the input – numerical age of the user and using the rule base created earlier will determine the category (group) of the users and which content should be submitted.

4 The defuzzification process

At this stage, the linguistic variable content will indicate to system from which group should be the content presented. If the content of the website is intended for middle age people, it will be very useful (maximum usefulness will be 1) for the 34-44 aged users. For users which age is 24-34 and 44-54 it will have a medium usefulness (0.2 and

0.8) and for the users younger than 24 years and older than 54 years, this content will have a lower usefulness, so will present no interest. For this reason the maximization function will be used [4]

$optimalContent = Max(content)$

because the aim is to present the most relevant content for the corresponding user age. So, if the value of the affiliation function would be 1, the best suitable content would be presented.

5 Conclusion

The proposed method allows the automatic selection of the most optimal and suitable content to be presented to the website users. The main advantage is that if the user is 25-34 years old (the affiliation function of G_M will have the maximal value – 1), and if the content for this group of user is available, it will be displayed. Otherwise the best suitable content will be displayed.

References

- [1] *Google Analytics Feature List*, 2014. <http://google.ru/intl/analytics/features>.
- [2] C.A. Smith. *Automated continuous process control*. 2002, pp. 115–122.
- [3] L.A. Zadeh. *Fuzzy sets, Information and Control.*, vol. 8, 1965, pp. 338–353.
- [4] L.A. Zadeh. *Computing With Words. Principal Concepts and Ideas.*, 2012, pp. 21–24.

Sergiu Chilat

Technical University of Moldova
Email: chilatsergiu@gmail.com