Nierówności Społeczne a Wzrost Gospodarczy, nr 44 (4/2015), część 1
DOI: 10.15584/nsawg.2015.4.1.3 **ISSN 1898-5084**

dr Germanas Budnikas[1]

Faculty of Economics and Informatics Vilnius branch
University of Bialystok

Creation of User's Online Behaviour Analysis Model for Increase of an Enterprise Competitiveness

INTRODUCTION

Currently World Wide Web resources (including various databases) are highly used by an extremely large amount of customers. They have an impact on the general use of e-services. In [*The social*, 2012] it is predicted from 0.9 to 1.3 trillion dollars in annual value could be earned if social media is fully exploited. Social media are a part of global online services where users behave in some manner – they watch information put there and finalise their transaction with a web site in a certain way, e.g. pay for services, comment posts, press "Like" button on Facebook and so on. In order to effectively provide business supported by the web site, it is of particular importance to understand user actions and interactions with a web portal – whether e-guests act in an appropriate manner; can we influence their online activities? This knowledge can be useful in behavioural targeting activities [Beales, 2014; Chen, 2014, pp. 429–449], in web site design [Cyr, 2014, pp 1–26] or in web site and customer interaction in order to facilitate on-line transaction finalisation. The last issue is described in the paper.

Similar research is conducted where user's on-line behaviour is analysed, classified into certain groups and forecasted (e.g. [Angeletou, pp. 35–50; Dembczyński, 2009, pp 189–206; Robinson, 2008, pp 100-109]). However, no works, which discussed attempts to influence actual on-line user behaviour, were found at present.

The object of the research is user's behaviour on web sites as well as its monitoring. The goal of the research is the creation of a user interaction model describing a process of user behaviour on a web site. The model's primary aim is to define the factors affecting a finalisation of user online transaction. Complimentary goal is the analysis of actual user interactions with a web site in order to in-

[1] Adres korespondencyjny: Faculty of Economics and Informatics Vilnius Branch, University of Bialystok, Kalvariju G. 135, LT-08221 VILNIUS, e-mail: german.budnik@uwb.edu.pl.

fluence his/her actions in case of approval a hypothesis about a possibility to not finalise the transaction.

Achieving the objective consists of three main steps: The first stage includes a formalization of behaviour on websites and determine the data needed for its description. This data will be read off from the web at the time of the user's interaction with a web site. At the second stage, an artificial neuron is created and taught, so that it could predict whether actual on-line user finalises the transaction or not based on the historical data of user behaviour. Developed model allows determining the most important factors affecting the finalisation of the transactions. The third stage consists of the analysis of actual user's behaviour by using the created model of the artificial neuron, during which it is determined whether the user has a tendency to finalise the transaction or not. If not, the portal may advice to visit those pages that have the greatest impact on finalisation of the transaction. All this activities have the effect on an increase of enterprise competitiveness.

The paper is organised in the following way – th*e motivati*on section discusses needs for such studies, its practical value as a factor of enterprise competitiveness. Succeeding section entitled *technique proposed* consists of subsections *metadata, exploring metadata* and *actual user behaviour data and its analysis. Metadata* subsection presents a generalised view of user behaviour on a selected category of a web site using Business Process Modeling Notation. Later the behaviour is formalised and its use is discussed in the subsection. A classification problem related to the on-line user behaviour is defined in subsection *exploring metadata*. A technique for solution of the classification problem – artificial neuron as well as requirements for its input data set are discussed there, too. Subsection *actual user's behaviour data and its analysis* describes how aforementioned techniques work together to influence on-line user behaviour and thus an overall competitiveness of an enterprise. The proposed model is compared with other ones in *related works* section. Conclusions summarize the paper.

MOTIVATION

Nowadays, a must to be part of a modern e- or m- commerce is a supporting web site. Its effective functioning during information supplying as well as during on-site business transactions can be regarded as a factor of an enterprise competitiveness. Therefore, it is very important to ensure that the web site gives as much as possible profit.

One of the emerging areas in recent years is consumer behaviour tracking on web sites. Through the Internet, business obtains additional source of information that can be employed. "The information is then cleaned, organized, and analysed using a number of statistical and data mining techniques to create a "shopping" profile of that individual. These profiles can then be used to target ad campaigns, personalize a shopping experience, or make recommendations on additional products" [Robinson, 2008, 100–109].

The idea of the paper has arose from a real situation of user's online behaviour. Most of commercial web sites supply various information on a product or a service to engage a visitor to make an order, that is, to finalise their transaction with a web site. Due to various reasons a part of visitors not finalise their transactions. It is believed that if a web site could estimate visitor intentions to not finalise their transaction, and afterwards, could interact with web visitor by suggesting a set of actions leading to increase of interest to order the item, then it would stipulate the increase in the number of users who finalise their transactions. That in its turn will affect enterprise competitiveness.

In order to implement this idea, a corresponding model should be created. It shall use statistical data on user's interaction with web site in the past and actual user's online interaction. Collecting of these data is possible due to Event Tracking method (a part of Google Analytics Tracking Code [Clark, 2014, pp 185–194]) that enables recording user interaction with website elements, such as embedded AJAX page element, page gadgets, Flash-driven element and so on.

Next section will describe model creation and employing phases in detail.

TECHNIQUE PROPOSED

Metadata

Due to a vast variety of different sorts of web pages, many different behaviour models could exist. To make model creation possible a kind of categorisation should be used to be applied to web sites. This can be done by employing the results of Open Directory Project [Bennett, 2012, pp 185–194; White, 2013, pp 1411–1420] that is an open content directory of web links constructed and categorised by humans. For illustration purposes, "Consumer Goods and Services" has been selected as a category of Open Directory Project. General user behaviour was outlined based on the analysis of several web sites of the selected category. Analysis results are depicted in Figure 1 using Business Process Modeling Notation (BPMN) [Drejewicz, 2012, p. 133].

According to the figure 1, a transaction with a web site of the selected category consists of the tasks depicted – product category selection, product selection and, next, viewing a related information about a product or service, that could be followed by viewing information about company, service or product delivery and making the order, that is, check-out and payment.

This task is considered as a final step in a whole transaction between a visitor and a web site and is referenced as a transaction finalisation. Note that every task corresponds to a web site page or its part. Thus, two general classes of user online behaviours can be derived – 1) visitors who finalise web site transaction during their Internet session, and 2) those who not finalise their transaction. Obviously different users will behave online in different ways that can be formalised by a quadruple:

$$\langle u, t, d, m \rangle$$

where:

u – *(user)* is an IP (internet protocol) address through which a user connects to a web site (note that several persons who use the same IP address are considered as a single user);

t – *(task)* is a kind of activity performed by a user on an Internet site page like "Product Category Selection", "Viewing Product Price Comparison" (see Figure 1);

d – *(duration)* is a time duration by which user performs a task;

m – is *moment of time*.

The defined quadruple sets requirements for:

1. Information content to be gathered from a web site by means of Google Analytics TrackEvent() function, and
2. Database table structure where information about user's interaction with a web page should be stored for further processing.

Google Analytics function TrackEvent() added to each web site page enables behavioural data sending from user computer and saving them in a database. These data are considered as metadata needed in further analysis steps.

Exploring metadata

Next, behaviours by individual users are important to investigate for discovering the trend (it is supposed that it exists) – what people do and how they behave in case they finalise their transactions and contrariwise.

Following this, two models defining the most influencing factors of user online behaviour with respect to metadata should be created:

1. model for behaviour leading to transaction finalisation, and
2. model for behaviour not leading to transaction finalisation.

Metadata gathered and stored in the database can be regarded as a relation *R*:

$$R: \langle u, t, D \rangle \rightarrow C_i$$

where:

D – is a sum of all durations *d* a user performs the same task *t* for all time moments *m*;

C_1 – corresponds a class – *transaction has been finalised*;

C_2 – corresponds a class – *transaction has not been finalised*.

That corresponds to a classification problem given an input data set. Artificial Neural Networks (ANN) can solve such kind of problems [Russell, 1995]. Namely, a Perceptron model using a supervised learning can classify input data into two classes – C_1 and C_2.

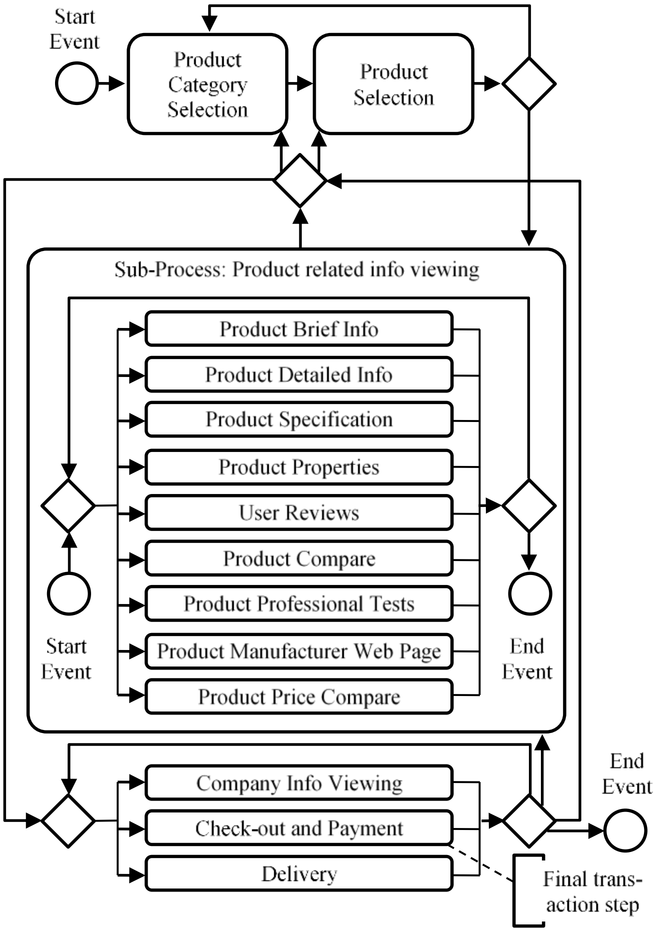

Figure 1. Generalized view of user behaviour on "Consumer Goods and Services" sites using BPMN notation

Source: Self-made.

Next, general parameters of the developed Perceptron model for classification of user's online behaviour are discussed. As Perceptron inputs should be in the interval [0,1] the following assumptions are made:

$t_i = {}^{1}/_{13} \cdot i$, where i is the number of the task and 13 is a total number of tasks as depicted on Figure 1 (the final task "Check-out and Payment is excluded);

$d_i = {}^{1}/_{3600} \cdot secs$, where *sec*s is a total time in seconds user surfs a web page during all sessions with a web site page for a given task t_i; 3600 seconds is set as a maximum surfing time for a given task during all sessions. Unvisited i-th web site page is denoted by $d_i = 0$.

In order to create well-learned Perceptron model, an artificial neuron error level should be set low, e.g. 0.05. It will impose a requirement for a number of training data set. According to a rule of thumb, its number should be equal to at least 520 (number of weights / error level = (13+13) / 0.05 = 520).

A kind of activation function as well as structure of a hidden layer is defined during experiments with a data set.

ANN related software like MATLAB [Demuth, 1993], Neuroph Studio [*The social*, 2014], etc. can be used to teach artificial neuron that will be used for classification of user's online behaviour. Most distanced weight values $w_i^{t|d}$ from a mid-value will indicate a list of corresponding task impact levels on a finalisation of user online behaviour. This list is used in definition of two values to be used later*:* most impactable tasks (MIT for short) and number of most impactable tasks (NMIT).

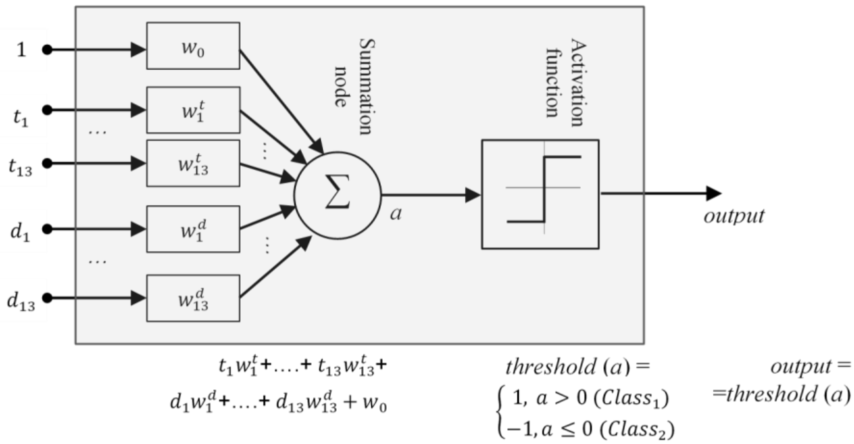

Figure 2. Perceptron model for classification of user online behaviour

Source: Self-made.

When Perceptron model is trained with aforementioned number of data, it can be used for evaluation of user's actual online behaviour. This procedure is explained in the succeeding section.

Actual user's behaviour data and its analysis

By monitoring user's online behaviour it is possible to detect a deviation of actual user's behaviour with respect to the desired one, that is one, which leads to finalisation of transaction as was defined in the previous section. After detecting such fact, actions to influence actual user's behaviour can be undertaken, e.g. online expert system could offer to visit web pages most important in finalising the transaction or even propose a discount. These issues will be discussed next.

During user actual browsing on a web site, Google Analytics TrackEvent() function calls enables collecting the behaviour data needed for analysis whether web site guest tends to finalise a transaction or not.

When distinct number of visited web pages, i.e. number of accomplished tasks t_i, reaches NMIT value, accumulated TrackEvent() values about actual user online behaviour are passed to the input of a trained Perceptron model (see the previous section). The Perceptron model will estimate a class of user online behaviour – web site transaction will be finalised or not. In a later case, not-started tasks with respect to the MIT list can be offered to perform in order to affect actual user online behaviour.

In case of all tasks from the MIT list have been executed, several actions could be undertaken according to an enterprise business policy that can be implemented in a web site expert system. An interaction loop between a user and a web site during which most impactable web pages are offered to visit should last till user finalises a transaction or leaves the web site.

Actual user's behaviour data also are accumulated in metadata database in order to periodically re-start Perceptron teaching with new data.

RELATED WORKS

The topic of the paper has quite important practical value. Analysis and understanding of web user behaviour is a key topic of behavioural targeting. Behavioural targeting is an evolving area of web mining that deals with optimisation of web online ads based on analysis of web user's behaviours.

The model presented in the paper has some similarities to works in the considered field of study and is characterised by few differences.

Methods of behavioural analysis investigate web surfing data mainly gathered from log files. The topic is actively investigated; examples of similar works could be papers by Angeletou *et al* [Angeletou, 2011], Dembczyński *et al* [Dembczyński, 2009], Robinson *et al* [Robinson, 2008], Ugtschmidt [Ugtschmidt, 2013], Xian *et al* [Xian, 2014, pp 34–47].

Robinson *et al* approach suggests a method for monitoring user online behaviour. The method is implemented based on data pulled from log files where HTTP/GET requests are saved when user clicks a hyperlink. These data are gathered using agent devices installed on user computers. The approach uses Open Directory Project for visited web site categorisation. The research emphasises on creation of behaviour profiles with respect to web page visitation event; frequencies and probability distributions; and causality relations or time-dependencies.

Angeletou *et al* [Angeletou, 2011] approach presents an approach of modelling and analysis of user behaviour in online communities that include person profiles, wiki, blogs, file sharing and forum. The approach implements behaviour modelling, role mining and role inference and is based on a statistical clustering.

Dembczyński *et al* [Dembczyński, 2009] technique describes the problem of predicting behaviour of web users based on real historical data. The data are gathered from user cookie files. The analysis is performed using statistical decision theory.

Xian *et al* [16] created a model and a technique to predict undesirable network behaviours, to find a relationship between network services and types of services provided, to define ratio of number of students employing typical services and flow ratio, to detect trends and types of services provided in the same period. Author used data mining, data warehouse and statistical analysis techniques. The following elements were monitored: specific patterns and rules in network user behaviour; time of using campus network; times and trends of using types of services provided; network use in different periods. The following data sources were used during 3 week monitoring activities: web site classification based on Open Directory Project (ODP); ODP catalogue; information from calendar, class and institution, user's IP address; log data of router; Domain Name Server catalogue; and e-card log data.

Ugtschmidt [Ugtschmidt, 2013] developed a model and tool for monitor user behaviour on media sites to detect behavioural differences between the task categories. Author uses the following techniques: descriptive analysis, mean value comparison and correlation analysis, testing hypotheses with mean value comparisons, machine learning methods, descriptive statistics of behavioural attributes, discriminant analysis. Accomplished experiment was performed on single site. During 4 weeks behaviours of 45 users were tracked using special software. Additionally users documented a kind of task they implemented during browsing – fact finding, information gathering, just browsing.

The approach proposed differs from works listed by its application area, i.e. in the Internet, while Angeletou *et al* [Angeletou, 2011], Robinson *et al* [Robinson, 2008] and [Xian, 2014] approaches operate at the Intranet level.

The approach proposed is similar to Angeletou *et al* [Angeletou, 2011] because they both uses dynamical update of estimations with respect to new data. The approach suggested and [Xian, 2014] are similar in a view of used source data – IP address and ODP catalogue. Ugtschmidt [Ugtschmidt, 2013] and the presented in the paper approach use machine learning method for analysis of gathered users' on-line behaviour data.

Although the approach presented in the paper suggests a model only and no practical experiment was given, it is possible to emphasise the following key differences:

– the model incorporates a technique for increase of enterprise competitiveness,
– most impactable factors on web site goal achievement can be defined using mathematical model based on actual user browsing data.

CONCLUSION

The goal of the research was creation of user's online behaviour analysis model. The goal was achieved by accomplishing a set of phases. A general user's behaviour scenario at a web page of category "Consumer Goods and Services" with respect to Open Directory Project was formalised using Business Process Modeling Notation. The data needed to represent a behaviour were defined and include user's IP address, task identifier and time moment to be passed using Google Analytics TrackEvent() function while user browses a web page. Historical data about user's behaviour are used for an artificial neuron training. After the training, most important factors that affect transaction finalisation (e.g. service order) can be defined.

While monitoring actual online user's behaviour the developed artificial neuron model permits to estimate behaviour outcome – would it be finalised or not. In a later case, it is foreseen an interaction of a web site, which can suggest visiting pages that has the highest impact level for transaction finalisation. In such a way, competitiveness of an enterprise is achieved. Future works in this direction include performing practical experiments and applying statistical techniques for analysis of user behaviour data. This will permit to define and investigate the most effective approach for user online behaviour analysis model.

BIBLIOGRAFIA

Angeletou S., Rowe M., Alani H., 2011, *Modelling and Analysis of User Behaviour in On-line Communities*. The Semantic Web – ISWC 2011, Lecture Notes in Computer Science Volume 7031, http://dx.doi.org/10.1007/978-3-642-25073-6_3.

Beales H., *The value of behavioral targeting*, Network Advertising Initiative, http://www.turn.com/sites/default/files/wp-content/uploads/2010/06/Beales_NAI_Study.pdf, Retrieved on, 1-09-2014.

Bennett P.N., White R.W., Chu W., Dumais S.T., Bailey P., Borisyuk F., Cui X., 2012, *Modeling the impact of short- and long-term behavior on search personalization*, proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR'12, http://dx.doi.org/10.1145/2348283.2348312.

Chen J., Stallaert J., 2014, *An economic analysis of on-line advertising using behavioral targeting*, Mis Quarterly, 38(2), http://dx.doi.org/10.2139/ssrn.1787608.

Cyr D., 2014, *Return visits: a review of how Web site design can engender visitor loyalty*. Journal of Information Technology, 29(1), pp 1–26, http://dx.doi.org/10.1057/jit.2013.25.

Clark D.J., Nicholas D., Jamali H.R., 2014, *Evaluating information seeking and use in the changing virtual world: the emerging role of Google Analytics*, Learned Publishing, 27(3), http://dx.doi.org/10.1087/20140304.

Dembczyński K., Kotłowski W., Sydow M., 2009, *Effective Prediction of Web User Behaviour with User-Level Models*, Journal Fundamenta Informaticae, 89(2-3).

Demuth H., Beale M., 1993, *Neural network toolbox for use with MATLAB*.

Drejewicz S., 2012, *Zrozumieć BPMN. Modelowanie procesów biznesowych*, Helion.

Robinson D.J., Berk V.H., Cybenko G.V., 2008, *On-line Behavioural Analysis and Modeling Methodology* (*OBAMM*), Social Computing, Behavioural Modeling, and Prediction.

Russell S., Norvig P., 1995, *Artificial Intelligence. A modern approach*, Prentice-Hall, Egnlewood Cliffs.

Sevarac Z., Goloskokovic I., Tait J., Morgan A., Carter-Greaves L., 2008, *Neuroph main page: resource for Neuroph*, Retrieved on, 13-09-2014.

The social economy: Unlocking value and productivity through social technologies, 2012, McKinsey Global Institute.

Ugtschmidt A., 2013, Classification of User Tasks by the User Behavior: Empirical Studies on the Usage of On-Line Newspapers, Berlin GmbH: Logos Verlag.

White R.W., Chu W., Hassan A., He X., Song Y., Wang H., 2013, *Enhancing personalized search by mining and modeling task behavior*, proceedings of the 22nd International Conference on World Wide Web, ACM WWW'13, pp 1411–1420.

Xian X., Chen F. and Wang J., 2014, *An Insight into Campus Network User Behavior Analysis Decision System*, Taichung, http://dx.doi.org/10.1109/IS3C.2014.145.

Summary

The goal of the research was to create a user's online behaviour analysis model. The goal was achieved by accomplishing a set of phases. A general user's behaviour scenario at a web page of category „Consumer Goods and Services" with respect to Open Directory Project was formalised using Business Process Modeling Notation.

Needed data to represent behaviour was defined and include user's IP address, task id and time moment to be passed using Google Analytics TrackEvent() function while user browses a web page. Historical data about user behaviour are used for an artificial neuron training. After the training, most important factors that affect transaction finalisation (e.g. service order) can be defined. While monitoring actual on-line user's behaviour the developed artificial neuron model permits the estimation of a behaviour outcome – would it be finalised or not. In a later case, it is foreseen an interaction of a web site, which can suggest visiting pages that has the most impact level for transaction finalisation. In such a way, competitiveness of an enterprise is achieved.

Keywords: online behaviour tracking, artificial neuron, Google Analytics, transaction finalization

Budowa modelu analizy zachowań użytkownika w Internecie dla wzrostu konkurencyjności przedsiębiorstw

Streszczenie

Celem badania jest stworzenie modelu analizy zachowania użytkowników w Internecie. Cel został osiągnięty poprzez wykonanie kilku czynności. Przede wszystkim został stworzony uogólniony scenariusz zachowania użytkownika na stronach typu *Consumer Goods and Services*, który został określony w *Open Directory Project,* używając notacji *Business Process Modelling Notation.*

Określone zostały również dane, odzwierciedlające zachowania użytkownika, które zawierają IP adres użytkownika, identyfikator strony WWW oraz czas. Te dane będą przesyłane używając funkcji *Google Analytics TrackEvent* w czasie, kiedy użytkownik ogląda stronę WWW. Dane historyczne o zachowaniu online użytkowników są używane dla uczenia neuronu sztucznego. Po uczeniu mogą być określone najbardziej skuteczne czynniki, wpływające na finalizację transakcji (np. zamówienie usługi). W czasie monitorowania faktycznego zachowania online użytkownika stworzony sztuczny neuron pozwala prognozować wynik oglądania stron WWW – czy transakcja będzie finalizowana lub nie. W przypadku zagrożenia niefinalizowania transakcji – witryna internetowa może sugerować odwiedzenie stron WWW, mających największy wpływ na przyjęcie decyzji o finalizowaniu transakcji. W ten sposób może być osiągnięte zwiększenie konkurencyjności przedsiębiorstwa.

Słowa kluczowe: śledzenie zachowania w Internecie, sztuczny neuron, Google Analytics, finalizacja transakcji

JEL: H00, C45, C59