Expert Systems with Applications 39 (2012) 13026–13034

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Predicting e-commerce company success by mining the text of its publicly-accessible website

Dirk Thorleuchter [a,*], Dirk Van den Poel [b]

[a] Fraunhofer INT, D-53879 Euskirchen, Appelsgarten 2, Germany
[b] Ghent University, Faculty of Economics and Business Administration, B-9000 Gent, Tweekerkenstraat 2, Belgium

ARTICLE INFO

Keywords:
Success factor
E-commerce
LSI
Classification
Text mining
Website

ABSTRACT

We analyze the impact of textual information from e-commerce companies' websites on their commercial success. The textual information is extracted from web content of e-commerce companies divided into the Top 100 worldwide most successful companies and into the Top 101 to 500 worldwide most successful companies. It is shown that latent semantic concepts extracted from the analysis of textual information can be adopted as success factors for a Top 100 e-commerce company classification. This contributes to the existing literature concerning e-commerce success factors. As evaluation, a regression model based on these concepts is built that is successful in predicting the commercial success of the Top 100 companies. These findings are valuable for e-commerce websites creation.

1. Introduction

Investigating the impact of information systems on the commercial success is a well-known topic by many researchers and practitioners (Ballantine, Levy, & Powell, 1998; Delone & McLean, 1992; Irani & Love, 2002; Lee & Kozar, 2006; Themistocleous, Irani, & Love, 2004). This impact only can be measured indirectly (Galletta & Lederer, 1989) because information systems are cross-linked socio-technical entities (Serafeimidis & Smithson, 2003) with intangible benefits and indirect costs (Irani, 2002). Literature focuses this challenging task by proposing success factors that impact information system's success (Delone & McLean, 1992; Serafeimidis & Smithson, 2003). E-commerce is a specific line of business and it differs from other line of business by the fact that the success of e-commerce companies strongly depends on companies' website quality (Carnero, 2005; Lee & Kozar, 2006; Lohse & Spiller, 1999; Ngai, 2003). Thus, specific success factors for e-commerce companies are necessary for considering these website quality aspects.

In literature, many e-commerce success factors for e-commerce companies are described (Baecke & Van den Poel, 2010). They are used to evaluate website quality (Zvirana, Glezerb, & Avnia, 2006) and thus to evaluate company's success. Examples for these factors are the usability of the web page, a human computer interaction, a well-known brand, a price reduction, and a money-back guarantee. The occurrence of these factors on the company's website can be used to predict the success of the e-commerce company (Thorleuchter & Van den Poel, 2012c).

Success factors are described on companies' websites in form of semantic textual patterns and they can be identified by searching for these patterns on the website content. As example, the occurrence of a textual pattern: 'A refund will be made if you are not satisfied.' or the occurrence of a textual pattern: 'We will return your money within the first 90 days' shows that the company offers a money-back guarantee. Although both textual patterns are formulated in different ways by use of different words, they share the aspect of meaning. A semantic textual pattern includes all textual patterns with the same meaning and thus, it can be used to represent the aspect that the company offers a money-back guarantee. Further examples are that the occurrence of semantic textual patterns describing the aspect of website usage in detail or describing the aspect of interactive website services shows that the company's website is of high usability or that a human computer interaction is implemented. The occurrence of semantic textual patterns that lay great stress upon company's name or a product's name (e.g. evoked by a high frequently occurrence of specific terms) on a company's website gives a hint that company's or product's name is probably a well-known brand. As shown from these examples, the use of e-commerce success factors by a company can be identified by extracting and analyzing semantic textual patterns from company's website. However, a manual analyzing of these patterns (e.g. by human experts) from the websites of many e-commerce companies is time-consuming (Thorleuchter & Van den Poel, 2012b).

* Corresponding author. Tel.: +49 2251 18305; fax: +49 2251 18 38 305.
E-mail addresses: Dirk.Thorleuchter@int.fraunhofer.de (D. Thorleuchter), dirk.vandenpoel@ugent.be (D. Van den Poel).
URL: http://www.crm.UGent.be (D. Van den Poel).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.05.096

Latent semantic indexing (Christidis, Mentzas, & Apostolou, 2012; Kim, Choi, & Kim, 2012; Lee & Wang, 2012; Shi & Setchi, 2012; Tsai, 2012) as an automated approach can be used in this case. In contrast to further text mining approaches, it also considers semantic aspects of the texts. The approach calculates several dimensions each representing a semantic textual pattern that occurs on the websites of the e-commerce companies. Each calculated pattern can be analyzed further by comparing its aspect of meaning to the aspects of meaning of the e-commerce success factors. As a result, semantic textual patterns can be identified that represent e-commerce success factors. Thus, an automated identification of semantic textual patterns from e-commerce companies' websites representing e-commerce success factors is possible by use of latent semantic indexing.

In contrast to the existing literature concerning e-commerce success factors, we contribute two new aspects to the scientific community. The first aspect as already described above is to identify e-commerce success factors that are used by e-commerce companies based on their semantic textual patterns with latent semantic indexing. The second aspect is to evaluate the successfulness of these factors by use of a semantic textual pattern based logistic regression as modeling technique (Coussement & Van den Poel, 2008). For this second task, we extract the textual content published on the websites of the Top 100 successful e-commerce companies and of the Top 101 to 500 successful e-commerce companies separately. The identified semantic textual patterns are evaluated in terms of their usefulness for predicting most successful e-commerce companies (Top 100) in contrast to less successful e-commerce companies (Top 101 to 500). Whereas the semantic textual patterns represent success factors, this means that existing success factors for predicting successful e-commerce companies are analyzed to show their success or to show their non-success in predicting most successful e-commerce companies. These results give useful insights for e-commerce decision makers, they contribute to the existing e-commerce success factor literature, and they are valuable for e-commerce websites creation.

This work uses web mining (Thorleuchter, Van den Poel, & Prinzie, 2010c) for crawling textual information from the Top 500 e-commerce companies' websites where a combined web structure mining and web content mining approach is processed. The web structure mining approach is adapted to the identification of e-commerce success factors. As an example, aspects of trustfulness (Thorleuchter & Van den Poel, 2011c; Thorleuchter, Weck, & Van den Poel, 2012a; Thorleuchter, Weck, & Van den Poel, 2012b) are considered by crawling textual information from web pages that contain information about the awarded certifications.

In sum, the provided methodology enables the prediction of e-commerce companies' success based on information extracted from the website content of e-commerce companies. The findings in this paper give valuable insights to decision makers of e-commerce companies by identifying success factors and by evaluating the impact of the success factors on company's success. These findings also are valuable for e-commerce websites creation and they contribute to the existing e-commerce success factor literature.

2. Background

2.1. Success factors for information systems

Companies have done much investment in the procurement, implementation, and processing of information systems. These systems should increase the productivity, improve the competitiveness, and reduce operational as well as administrative costs (Molla & Licker, 2001; Schuette, 2000). Thus, decision makers of companies are interested in evaluating the success of the information systems e.g. to calculate the return on investment. Based on finding of (Galletta & Lederer, 1989) that the success of information systems can only be measured indirectly, literature shows two different ways for evaluating the success of information systems. On one hand information systems and their impact on company's success are modeled and based on the modeling results, new approaches are proposed for the evaluation (Irani, 2002; Irani & Love, 2002; Mcaulay, Doherty, & Keval, 2002; Smithson & Hirschheim, 1998). On the other hand, success factors are identified that impact information system's success (Delone & McLean, 1992; Serafeimidis & Smithson, 2003).

A well-known information system success model that is based on success factors is from Delone and McLean (1992). The model also contains the interdependencies of the success factors and it leads to three conclusions: First, the quality of the information as well as the quality of the system itself impacts its success. Second, information systems that are easy to use and that consists of a high user satisfaction also are successful. Last, information systems with high impact on individuals and on organizational structures are more successful than other.

2.2. Success factors for e-commerce

Measuring the success of information systems by applying information system success factors is interesting for decision makers of companies. E-commerce is a specific line of business and e-commerce decision makers have done much investment in information systems in particular in their e-commerce website and they are also interested in evaluation their information systems. Literature has shown that the success of an e-commerce company strongly depends on the quality of its website (Carnero, 2005; Lee & Kozar, 2006; Lohse & Spiller, 1999; Ngai, 2003). Thus to improve commercial success, e-commerce decision makers have the possibility to improve their website quality by considering e-commerce success factors. Specifically for e-commerce, much work has been done by researchers to identify these factors (Baecke & Van den Poel, 2011; DeBock & Van den Poel, 2009; Delone & McLean, 1992; Lee & Kozar, 2006; Lopeza & Ruiz, 2010; Lu, Zhao, & Wang, 2010; Serrano-Cinca, Fuertes-Callén, & Gutiérrez-Nieto, 2010; Van den Poel & Buckinx, 2005; Verhoef et al., 2010).

McKinney, Yoon, and Zahedi (2002) transfer success factors from Delone and McLean (1992) to the e-commerce domain. A high quality of the used information and a high quality of the website system itself lead to an increased internet customer satisfaction and thus, to commercial success. Devaraj, Fan, and Kohli (2002) found that the e-commerce success is increased if customers found useful information on the website and the website itself is easy to use. Further factors are that the responding time of the website is low and that price savings are offered to the customers. Torkzadeh and Dhillon (2002) show that a wide internet product choice and that the online payment are factors that lead to an increased internet shopping convenience. Further, the internet customer relation can be improved if customers identify an internet vendor as trustful. Further factors are the shopping travel and the internet shipping that have an impact on internet shopping convenience, internet ecology, and internet product value. Unfortunately, no validation of the impact of internet shopping convenience, internet ecology, internet customer relation, and internet product value on e-commerce success is given by Torkzadeh and Dhillon (2002). Zhu and Kraemer (2002) show e-commerce success factors based on an evaluation on 260 manufacturing companies. Besides the already mentioned success factor 'information quality', an easy processing of the purchase also increases e-commerce success. A further factor is the customization of the website where a website is presented to each customer in an individual way. Additionally, e-commerce success can be increased by a direct website based supplier connection where goods can be delivered just in time.