Expert Systems with Applications 41 (2014) 5009–5016

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Semantic weak signal tracing

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

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

ARTICLE INFO

Keywords:
Time series
Trend identification
Latent semantic indexing
Web mining

ABSTRACT

The weak signal concept according to Ansoff has the aim to advance strategic early warning. It enables to predict the appearance of events in advance that are relevant for an organization. An example is to predict the appearance of a new and relevant technology for a research organization. Existing approaches detect weak signals based on an environmental scanning procedure that considers textual information from the internet. This is because about 80% of all data in the internet are textual information. The texts are processed by a specific clustering approach where clusters that represent weak signals are identified. In contrast to these related approaches, we propose a new methodology that investigates a sequence of clusters measured at successive points in time. This enables to trace the development of weak signals over time and thus, it enables to identify relevant weak signal developments for organization's decision making in strategic early warning environment.

1. Introduction

Strategic planning for an organization can be improved by the identification and use of signals (Ansoff, 1975). A signal is defined as an event with impact on a specific target or direction. While strategic planning aims on defining directions, signals have to be considered during strategic decision making process. Weak signals are signals where the impact cannot be estimated accurately (Ansoff, 1984). It might be that a new event will possibly have an impact on a target in future. It also might be that an existing event – that does not have an impact on the target up to now – will possibly have an impact in future. For strategic planning, it is hard to identify weak signals from the large number of existing signals. Literature proposes methodologies for weak signal identification. They can be used to identify the future impact of weak signals on own strategic directions.

Weak signals cannot be found in the core area of an organization. This is because all internal events of an organization and their impacts normally are already known by strategic decision makers. Thus, Ansoff shows that weak signals can be found in organization's environment. This requires the use of an environmental scanning procedure to identify signals in a first step. This also requires the use of a clustering approach to group the large number of identified signals and to identify clusters of weak signals in a second step.

The concept of environmental scanning (Tonn, 2008) aims at extracting and analyzing information from various data sources existing in the organization's environment. After analyzing, events can be identified as well as their relationships. Today, the internet is a large and valuable source of information (Decker, Wagner, & Scholz, 2005) where many signals occur. Further, the internet can be used to represent organization's environment. Additionally, most of the data available in the internet are textual data, e.g. websites or blogs. As a result, existing weak signal identification approaches use an environmental scanning that considers textual information from the internet (Decker et al., 2005; Uskali, 2005).

With an internet based environmental scanning, documents e.g. webpages can be identified. This scanning normally has a wide scope and thus, it leads to a large number of extracted internet documents. This makes the use of a (semi-) automatic approach more appropriate than the use of a manual approach. The documents possibly contain texts related to several different topics. Thus, a document as a whole normally does not represent a signal but specific textual patterns that occur within the document probably do (Uskali, 2005). Text mining can be used to extract textual patterns from the full text of the documents and a specific clustering approach can be applied to identify groups of textual patterns that represent weak signals (Tabatabei, 2011; Thorleuchter & Van den Poel, 2013a).

Literature shows some approaches that use internet based environmental scanning for weak signal identification. The approach of Schwarz (2005) aims at the identification of new arising technologies with relevance for the high tech companies in Europe. Unfortunately, the approach could not be applied in practice. It has caused a very high manual effort because an automated environmental

* 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).

http://dx.doi.org/10.1016/j.eswa.2014.02.046
0957-4174/

scanning tool was not available and thus, the scanning was processed by human experts. Further, the results of the clustering approach are of low quality. In contrast to this, the approaches of Decker et al. (2005) and Uskali (2005) have been applied successfully. However, they prevent the high manual effort by restricting the number of retrieved documents to a small value. Thus, they could not be seen as wide scope internet based environmental scanning approaches. Tabatabei (2011) provides an automated approach for internet based environmental scanning and clustering. A further knowledge structure based approach is provided by Yoon (2012) that detects weak signal from internet news related to solar cells. Both approaches are applied in a case study however, they are not evaluated. Thorleuchter and Van den Poel (2013a) proposes a semantic clustering approach that can be used together with internet based environmental scanning to identify weak signals. The strength of this semantic approach is that it considers weak signals where the corresponding text patterns are written by different persons, in different writing styles, and in different contexts. Text patterns are recognized as similar if they share a common meaning even if they do not share common words. Based on a clustering of these text patterns, signals can be identified and distinguished in weak and strong signals. The authors prove the feasibility of the proposed approach by evaluating results of a case study.

In contrast to related work, we provide a methodology that investigates a sequence of clusters that represent weak signals. This enables to trace the development of weak signals over time. In a first step, an internet based environmental scanning is processed and semantic clustering is applied to identify signals. This first step is processed in accordance to the methodology proposed by Thorleuchter and Van den Poel (2013a). In a second step, the internet based environmental scanning is repeated at successive points in time. For each point in time, the scanning results are projected into the semantic space created by the clustering approach in the first step. This allows tracing the identified signals. Examples are the identification of weak signals that lose their impact on a target, weak signals that become strong signals with large impact on a target, or weak signals that do not change its impact over time. For clustering and classification, latent semantic indexing (LSI) is used. It enables the identification of semantic textual patterns from large document collection and it also enables clustering and the assignment of new documents to an existing semantic space.

In a case study, the proposed methodology is applied in the field of storages technologies for intermittent energy sources where the development of weak signals is traced. The aim of the case study is to show the general feasibility of the approach. For evaluation, hypotheses about future development of storages technologies were provided by a literature review of future studies. They are compared to the traced weak signals. As a result, the proposed methodology enables to identify future technological developments based on current internet information about energy storage technologies. This supports decision makers by their strategic decision making.

Overall, we propose a methodology that enables to trace the development of weak signals over time. It also considers aspects of meaning from the documents identified by an internet based environmental scanning. Tracing weak signals can support strategic decision making because events with impact on strategic aims or directions can be identified in advance. This allows decision makers to react ahead of time.

2. Background

2.1. Internet based environmental tracing

A huge amount of information can be found in the internet dealing with different topics. Using this information together with traditional information sources (e.g. organization internal databases) provides an added value for decision making (D'Haen, Van den Poel, & Thorleuchter, 2013). An example for using this information for organization's strategic planning is to collect and analyze information from the internet about organization's customers and competitive organizations (Teo & Choo, 2001). The huge amount of information available in the internet enforces the use of web mining approaches. This enables to collect information based on an automated process for scanning all relevant internet websites (Kosala & Blockeel, 2000). Web mining includes the identification of website's relevance for a specific topic and it also includes the process of reducing information from relevant websites (Velásquez, Dujovne, & L'Huillier, 2011). The identification of website's relevance is normally realized by using advanced programming interfaces (APIs) of internet search engines (Thorleuchter & Van den Poel, 2013b). To reduce information from websites, automated filtering algorithms are applied. Web mining approaches can be evaluated based on the performance measures in information retrieval: the precision and the recall. Considering all relevant websites helps to improve the recall measure in information retrieval and considering results of a high quality filtering helps to improve the precision measure. Some web mining approaches are applied at successive points in time to enable an environmental internet tracing. They discover current trends and relevant changes from the internet (Loh, Mane, & Srivastava, 2011). Thus, internet information is a valuable source for strategic decision making in an organization.

2.2. Identification of signals and signal tracing

A well-known concept for implementing an early warning system used in strategic planning is introduced by Ansoff (1975) that focusses on the identification of signals, specifically weak signals. Signals are defined as events, e.g. future trends, changes, or further emerging phenomena with a specific impact on a given target (Yoon, 2012). It could be distinguish between strong signals and weak signals. A strong signal impacts a target at present above a specific threshold and it is expected that this signal also will impact the target in future (Mendonça, Cardoso, & Caraça, 2012). In contrast to this, a weak signal has none or a small impact on a target at present but possibly, it will get an impact on the target in future (Tabatabei, 2011). Thus, the identification of weak signals makes it possible for decision makers to be aware of events in advance that will impact the decision in future (Kuosa, 2010). A further definition of weak signals describes them as unstructured information with low content value at present time that reflects e.g. aspects of an opportunity or a threat without aiming at a specific target (Mendonça, Pina e Cunha, Kaivo-oja, & Ruff, 2004). If the content information becomes more concrete by mention the impact of the opportunity or threat on a specific target then a weak signal has become a strong signal (Holopainen & Toivonen, 2012).

In the internet, many webpages can be found where strong signals are mentioned. This is because their impact on a specific target is already known and they are widely discussed on several websites, new articles, and internet blogs. Thus, strong signal with impact on a specific target occur high frequently in the internet. In contrast to this, weak signals occur low frequently in the internet because they lack a current impact on a target and thus, they are not attractive for discussion and seldom mentioned on websites, new articles, and blogs. However, it might be that a small number of authors recognize the future impact of a weak signal and describe it in the internet. These few documents are among the large amount of information available in the internet. The identification of these documents and thus, the identification of weak signals in the internet is difficult and many practical approaches fail because of this information retrieval problem (Schwarz, 2005).