

Chapter 86

Discovering Patterns in Order to Detect Weak Signals and Define New Strategies

Anass El Haddadi

University of Toulouse III, France & University of Mohamed V, Morocco

Bernard Dousset

University of Toulouse, France

Ilham Berrada

University of Mohamed V, Morocco

ABSTRACT

Competitive intelligence activities rely on collecting and analyzing data in order to discover patterns from data using sequence data mining. The discovered patterns are used to help decision-makers considering innovation and defining the strategy for their business. In this chapter we present four methods for discovering patterns in the competitive intelligence process: “correspondence analysis,” “multiple correspondence analysis,” “evolutionary graph,” and “multi-term method.”

INTRODUCTION

A successful business is often conditioned by its ability to identify, collect, process, and disseminate information for strategic purposes. However, a company can be over-informed, and not be able to search through all this information. Now, to be competitive it must know their environment. The establishment of a competitive intelligence (CI) approach is the inevitable answer to this challenge.

In the last few years, a lot of work has been done in order to ensure CI approaches. Discovering weak signals and define new strategies have been the main motivation for applying them in company contexts. The CI approach can provide the company with detailed information about its environment through internal and external information which it has access to. This environmental scanning is intended to assist decision makers in their choice of strategies.

DOI: 10.4018/978-1-4666-2455-9.ch086

In our CI approach, we use techniques for extracting knowledge from textual data to study scalable relational data from the information environment of a company. In this context, we propose our competitive intelligence tools: “TETRALOGIE” and “Xplor” (Web service of TETRALOGIE). These tools extract the weak signals and define new strategies using sequence data mining from a corpus. These patterns are used also in various areas: biology (Qindga and al., 2010; Shuang and Si-Xue, 2010), traffic prediction (Zhou *and al.*, 2009; Zhou *and al.*, 2010), space research (Walicki and Ferreira, 2010; Yun and al., 2008), and so on.

In this chapter, a CI approach based on sequence data mining is detailed. It uses four methods:

- **Correspondence Analysis (CA):** which aims at detecting the evolution of a research area, authors, company’s, keywords, etc... or the temporal sequence, that allows us to have an overview of changes in very specific areas.
- **Multiple Correspondence Analysis (MCA):** which aims at detecting the time series for decision making.
- **Multi-term Method:** which aims at extracting the weak signals.
- **Evolutionary Graph:** which shows in detail the structural changes of networks over time. For example, we detect the appearance and changes in social networks.

This document is organized as follows. First, we identify in section 1 the knowledge extraction process in order to demonstrate our methodology of analysis, and various measures of information structure. In section 2, we explain extraction of strategic information and the discovery of patterns by correspondence analysis (CA). Section 3 presents the patterns of weak signals, and describes methods to detect a pattern for new

strategies in a company. In Section 4, we explain the methodology to detect “temporal,” “pattern” sequences using evolutionary graphs. And finally in section 5, to illustrate the methods presented in the previous sections, a presentation of a complete analysis of emerging field of agronomy in China is performed by our research team.

KNOWLEDGE EXTRACTION PROCESS IN CI

The key step of the CI process is the selection of information, which is to develop a “corpus”, depending on the target, which will be later analyzed through methods of text mining. We often use the term “corpus” to describe large sets of semi or fully-structured textual data available electronically.

Following predefined criteria, this step allows us to focus on data defined as “interpretable” and with high informative potential. Data is firstly prepared by selecting it according to objectives fixed using the techniques of information retrieval (Büttcher *et al.*, 2010) (Croft *et al.*, 2010). This process (Saltan & McGill, 1984) seeks to match a collection of documents and the user needs (Maniez et Grolier, 1991), translated in the form of a request (Kleinberg, 1999) through an information system. This is composed of an automatic or semi-automatic indexing module, a module of document/request matching and possibly a module of query reformulation.

Different models are used in search engines to match the query with the document, such as the probabilistic model (Sparck-Jones, 2000), the connexionist and genetic model (Boughanem et al., 2000), the flexible model (Sauvagnat, 2005), the language modeling (Ponte & Croft, 1998), etc.

Monitoring devices are based on two types on information: formal and informal.

15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/discovering-patterns-order-detect-weak/73516

Related Content

Sentiment Analysis as a Restricted NLP Problem

Akshi Kumar and Divya Gupta (2022). *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines* (pp. 1032-1063).

www.irma-international.org/chapter/sentiment-analysis-as-a-restricted-nlp-problem/308532

Big Data Literacy: A New Dimension of Digital Divide, Barriers in Learning via Exploring “Big Data”

Dimitar Christozov and Stefka Toleva-Stoimenova (2016). *Big Data: Concepts, Methodologies, Tools, and Applications* (pp. 2300-2315).

www.irma-international.org/chapter/big-data-literacy/150266

Data Warehouse Testing

Matteo Golfarelli and Stefano Rizzi (2011). *International Journal of Data Warehousing and Mining* (pp. 26-43).

www.irma-international.org/article/data-warehouse-testing/53038

Classification and Machine Learning

Damian Alberto (2019). *Advanced Metaheuristic Methods in Big Data Retrieval and Analytics* (pp. 180-193).

www.irma-international.org/chapter/classification-and-machine-learning/216098

Preference-Based Frequent Pattern Mining

Moonjung Cho, Jian Pei, Haixun Wang and Wei Wang (2005). *International Journal of Data Warehousing and Mining* (pp. 56-77).

www.irma-international.org/article/preference-based-frequent-pattern-mining/1759