Using online textual data, principal component analysis and artificial neural networks to study business and innovation practices in technology-driven firms

Giacomo di Tollo [a], Stoyan Tanev [b,*], Giacomo Liotta [c], Davide De March [d,e]

[a] Dipartimento di Management, Università Ca' Foscari, San Giobbe, Cannaregio 873, 30121 Venezia, Italy
[b] Department of Technology and Innovation, Faculty of Engineering, University of Southern Denmark, Campusvej 55, Odense M, DK-5230, Denmark
[c] Center for Industrial Production, Department of Business and Management, Faculty of Engineering and Science, Aalborg University, A.C. Meyers Vaenge 15, DK-2450 Copenhagen SV, Denmark
[d] EvoSolutions S.r.l., Viale Ancona 17, 30174 Venice, Italy
[e] European Center for Living Technology, San Marco 2940, 30124 Venice, Italy

ABSTRACT

In this paper we introduce a method that combines principal component analysis, correlation analysis, K-means clustering and self organizing maps for the quantitative semantic analysis of textual data focusing on the relationship between firms' co-creation activities, the perception of their innovation and the articulation of the attributes of their product-enabled services. Principal component analysis was used to identify the components of firms' value co-creation activities and service value attributes; correlation analysis was used to examine the relationship between the degree of involvement in specific co-creation activities, the online articulation of firms' service value attributes and the perception of their innovativeness. K-means and self organizing map (SOM) are used to cluster firms with regards to their involvement in co-creation and new service development, and, additionally, as complementary tools for studying the relationship between co-creation and new service development.

The results show that, first, there is a statistically significant relationship between firms' degree of involvement in co-creation activities and the degree of articulation of their service value attributes; second, the relationship should be considered within the context of firms' innovation activities; third, OS Software-driven firms are the best example in terms of co-creation and new product-enabled service development, i.e. the collaborative principles built in their customer participation platforms should be adopted by other (non-software) firms interested in enhancing their innovation capacity through involvement in co-creation and new product-enabled service development.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

The recent years have witnessed an unprecedented interest in the conceptualization and development of research techniques and business intelligence tools using online textual data to generate insights about the business, marketing or innovation aspects of specific firms' or industry sectors. In some cases the focus is on data related to online communities (blogs, forums or review sites) and social networks; in other cases the focus is on company's online information which is expressed in natural language in order to be publicly disclosed to customers and other relevant stakeholders. Such information can be used to both analyze and inform specific aspects of firms' policy and strategic decision making with respect to customer relationship management, marketing communications, or open innovation management initiatives. In this article we aim to:

- demonstrate the value of public online textual data as a source of actionable information about specific product development, business and innovation practices in different industry sectors;
- validate a simple web search technique for textual data collection which is combined with principal component (PCA) to provide a systematic interpretive methodology that could be used in conceptualizing the relationship between different business and innovation practices;

- apply a combination of cluster analysis (CA) and an artificial neural network (ANN) approach using self-organizing maps (SOMs) as complementary methods enabling a more generic classification of firms in terms of the degree of their involvement in the specific business and innovation practices.

The analytical approach that was adopted in this article could be categorized as a modified and more intuitive version of latent semantic analysis (LSA) – a method for extracting the meaning from passages of text, based on a series of linear algebraic operations and statistical computations over a collection of documents or other textual sources such as e-mail messages or web pages [1–4]. The adoption of the LSA technique has visibly increased over the last 20 years. It is however a relatively complex approach with high barriers to adoption by the majority of scholars and practitioners. The analytical method suggested here aims at validating a more intuitive version of quantitative content analysis that could become more easily accessible through the use of open source software web search and computational tools as well as standard business statistics software packages such as SPSS, SAS, Microsoft Excel, etc. Unfortunately, user-friendly intelligent systems for business intelligence purposes are not widely available yet, thus entailing labor intensive and time consuming tasks for business analysts [5]. Our article addresses this issue by using a keyword-based web search on the online text of companies' websites to generate the number of occurrences of the keywords and a combination of well-known techniques such as PCA, correlation analysis, K-means CA and ANN to process the collected data and interpret it in a meaningful way. The set of keywords is preliminary constructed on the basis of subject domain expertise in a way that it could cover in a relatively complete way the various issues, aspects or activities associated with a particular business or innovation aspect. In this sense, there is a key difference between LSA and our approach to the selection of the keywords that are going to be used in the search process. In LSA, the initial set of keywords is usually coming from the documents themselves through a pre-processing step focusing on the identification of the terms that would be most relevant in the research process. In other words, LSA is looking to discover or uncover what is already in the documents without any preliminary focus on a specific theme or topic. In our approach we start with a specific theme of interest such as, for example, firms' involvement in co-creation or innovation practices, and then construct a set of keywords that is used as a probing tool or an exploratory lens to look into how the firms included in the research sample articulate their association with the particular theme or topic. The main sources of the initial set of keywords are research and practitioner publications focusing on the topic of interest, complemented by a parallel alignment of the terminology with selected websites of firms included in the research sample. One could say therefore that in our approach the set of keywords is external to the online text and not necessarily coming out of it. Such approach allows for some more flexibility in the construction of the logical combination of keywords including different words with similar or overlapping meanings. This is another important difference since it shows that the exploratory power of our approach is based on the business domain subject expertise and not so much on the expertise required for the development and the operation of the business intelligence tool.

A approach similar to the one used in this article was previously applied to the identification of the technology commercialization strategies of high-tech small firms [6] and the value co-creation practices of firms using open source (OS) technologies as part of their business model, showing that it allows for the classification of the firms in terms of the relationship between their co-creation and innovation activities [7,8]. Recently it was also applied to the identification of the product-enabled value attributes of top

Research and Development (R&D) spenders in Canada and Europe [9,10]. The present article represents a first attempt to apply the methodology to the simultaneous analysis of three different aspects – the degree of firms' involvement in value co-creation practices, the articulation of their product-enabled service value attributes and their own perception of the innovativeness of their products, processes and services. While the various innovation aspects of new and established firms have been traditionally the subject of significant interest, value co-creation practices and product services are in the process of attracting particular attention.

Value co-creation has emerged as a new business, marketing and innovation approach in which customer and end users are seen as active part in the design and shaping of personalized products, services and experiences [7,11–16]. It is based on the design and development of customer participation platforms, which provide firms with the technological and human resources, tools and mechanisms to benefit from the engagement experiences of individuals and communities [17–19]. The ability of such platforms to enable the personalization of new products and services challenges the operational presuppositions of traditional marketing segmentation techniques by promoting a new service-dominant logic [20] which allows firms to address broader heterogeneous markets aiming at a better fit between what customers need and what the firm makes and offers.

Product-enabled services or product-service systems (PSS) are considered as a way for firms' to be more innovative by offering more value to customers as well as to attain business differentiation and sustainability [21–24]. More specifically, PSS are usually associated with a specific type of value proposition consisting of a mix of tangible products and intangible services designed and combined in a way that they jointly are capable of fulfilling specific customer needs [23]. In this article the term product-enabled service will be used as an inclusive way of addressing the different ways of adding value by combining products and services within a specific business context. The application of the suggested methodology within the context of three different aspects demonstrates its ability to handle a variety of aspects related to business and innovation management.

The rest of the article is structured as follows. We outline our method is Section 2 and review the relevant literature on value co-creation practices and product-enabled services in Section 3. Results will be shown in Section 4 and discussed in Section 5. Finally, in Section 6 we summarize our concluding remarks and make suggestions for future research.

2. Methodology

The objective of this article is to examine the relationship between the degrees of firms' involvement in value co-creation practices, the articulation of the value attributes of their product-related services and the perception of their innovativeness. The study adopts a quantitative methodology which is based on: (i) online textual data that could be found on companies' websites; (ii) web search techniques for textual data collection combined with PCA to provide the key components of their value co-creation practices and the articulation of their service value attributes; (iii) cluster and ANN analysis to provide a generic classification of the firms in terms of the degree of their involvement in value co-creation and the degree of articulation of their service value attributes.

2.1. Research sample

The results are based on a sample of 230 firms that were selected for being representative of the breadth of their value co-creation activities. This is a subset of the research sample that