



Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation



Sérgio Moro^{a,*}, Paulo Cortez^b, Paulo Rita^c

^a Business Research Unit (UNIDE-IUL), Department of Information Science and Technology, ISCTE – University Institute of Lisbon, 1649-026 Lisbon, Portugal

^b ALGORITMI Research Centre, Department of Information Systems, University of Minho, 4800-058 Guimarães, Portugal

^c Business Research Unit (BRU-UNIDE), ISCTE – University Institute of Lisbon, 1649-026 Lisbon, Portugal

ARTICLE INFO

Article history:

Available online 29 September 2014

Keywords:

Banking
Business intelligence
Data mining
Text mining
Decision support systems

ABSTRACT

This paper analyzes recent literature in the search for trends in business intelligence applications for the banking industry. Searches were performed in relevant journals resulting in 219 articles published between 2002 and 2013. To analyze such a large number of manuscripts, text mining techniques were used in pursuit for relevant terms on both business intelligence and banking domains. Moreover, the latent Dirichlet allocation modeling was used in order to group articles in several relevant topics. The analysis was conducted using a dictionary of terms belonging to both banking and business intelligence domains. Such procedure allowed for the identification of relationships between terms and topics grouping articles, enabling to emerge hypotheses regarding research directions. To confirm such hypotheses, relevant articles were collected and scrutinized, allowing to validate the text mining procedure. The results show that credit in banking is clearly the main application trend, particularly predicting risk and thus supporting credit approval or denial. There is also a relevant interest in bankruptcy and fraud prediction. Customer retention seems to be associated, although weakly, with targeting, justifying bank offers to reduce churn. In addition, a large number of articles focused more on business intelligence techniques and its applications, using the banking industry just for evaluation, thus, not clearly acclaiming for benefits in the banking business. By identifying these current research topics, this study also highlights opportunities for future research.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Banking has been a prolific industry for innovation concerning information systems and technologies (Shu & Strassmann, 2005). For example, new technologies have enabled new communication channels which were quickly adopted by banks. Also, advanced data analysis techniques are currently used to evaluate risk in credit approval (Huang, Chen, Hsu, Chen, & Wu, 2004) and fraud detection (Ngai, Hu, Wong, Chen, & Sun, 2011).

Business intelligence (BI) is an umbrella term that includes architectures, tools, databases, applications and methodologies with the goal of analyzing data in order to support decisions of business managers (Turban, Sharda, & Delen, 2011). Banking domains, such as credit evaluation, branches performance, e-banking, customer segmentation and retention, are excellent fields for application of a wide variety of BI concepts and techniques, including

data mining (DM), data warehouses and decision support systems (DSS). For bank firms to survive and even excel in today's turbulent business environment, bank managers need to have a continuous focus on solving challenging problems and exploiting opportunities. That demands a need for computerized support of managerial decision making thus implying the need of decision support and business intelligence systems.

There are several surveys/reviews of the banking domain. Wilson, Casu, Girardone, and Molyneux (2010) published a recent literature review covering the impact of the global financial crisis in the banking business. Their results put the risk domain as a subject that deserves a deeper attention in order to achieve a systemic stability. The review of Ngai et al. (2011) devoted attention to financial fraud detection, and classified 49 articles depending on the type of fraud. The findings suggest a lack of research on mortgage fraud, money laundering, and securities and commodities fraud, by contrast to a large number of articles on credit fraud. More related with this paper, Fethi and Pasiouras (2010) presented a survey on bank branches performance based on 196 articles which employ operational research and artificial intelligence

* Corresponding author.

E-mail addresses: scmoro@gmail.com (S. Moro), pcortez@dsi.uminho.pt (P. Cortez), paulo.rita@iscte.pt (P. Rita).

techniques, concluding that profit efficiency and capacity efficiency have received limited attention in the studies evaluated.

A large research attention has been given toward credit. In fact, although credit is traditionally related to banking, it has long spread to other industries. Therefore, some recent reviews and surveys are naturally available on the subject. [Abdou and Pointon \(2011\)](#) reviewed 214 articles/books/thesis on credit scoring applications, searching for the statistical techniques used for evaluation and found that there is not an overall best technique for building models. The review of [Marqués, García, and Sánchez \(2012\)](#) reports over the use of evolutionary computation for credit scoring. Another subject of interest is e-banking, specifically customer acceptance toward a new communication channel. [Dahlberg, Mallat, Ondrus, and Zmijewska \(2008\)](#) reviewed publications on mobile payments and found through their framework lacking of research on social and cultural factors impacting mobile payments, as well as traditional payment services.

The enlisted surveys and reviews cover some themes in banking. However, within the authors' knowledge, there is a lack of a recent literature analysis for BI applications in the main subjects related to the banking industry, thus motivating the present research. Furthermore, none of the discussed reviews adopted an automated text analysis, by using Text Mining (TM) techniques such as the ones presented in this study, thus facilitating the analysis of a much larger set of sources.

This paper presents an automated text mining literature analysis, from 2002 to 2013, of BI applications within the banking domain, allowing the identification of current research trends and interesting future applications, thus highlighting opportunities for further research. Although BI has been extensively studied, recent years and particularly the last decade have experienced a huge increase in BI applications through the industry, especially in the banking sector, therefore stimulating research. This article is organized as follows. Section 2 introduces the main concepts related with both banking and BI domains, and presents also other references of literature analyses. Next, Section 3 presents the methods used for analyzing the literature. Then, the results are discussed in Section 4. Finally, conclusions are summarized in Section 5, which also presents future research directions.

2. Background

2.1. Text mining

Data mining (DM) aims to extract useful knowledge (e.g., patterns or trends) from raw data ([Witten & Frank, 2005](#)). Text mining (TM) is a particular type of DM that is focused on handling unstructured or semi structured data sets, such as text documents ([Fan, Wallace, Rich, & Zhang, 2006](#)). [Delen and Crossland \(2008\)](#) proposed the application of TM for analyzing the literature and identify research trends, thus helping researchers in conducting state of the art reviews on a given research subject. Their research focused on three major journals in management information systems, although they argue that their TM approach can be valuable in virtually any research field.

Within a literature analysis, searching with individual words is often not enough, since many searchable terms can be composed of a sequence of words, such as “data mining” or “decision support systems”. Those sequences, which can be made of n words, are called n -grams. When extracted from large texts, n -grams constitute a valuable asset, in particularly when analyzing publications, such as the study of [Soper and Turel \(2012\)](#) showed by analyzing eleven years (from 2000 to 2010) of publications in the Communications of the ACM journal.

When conducting TM over text documents, relevant words and terms are often extracted in order to produce a categorization that can help building a body of knowledge over the literature considered ([Delen & Crossland, 2008](#)). An interesting approach is modeling a certain number of distinct topics defined according to the number and distribution of terms across the documents, which can be achieved through the latent Dirichlet allocation (LDA) model ([Blei, Ng, & Jordan, 2003](#)). For each document, it is determined the probability of belonging to each of the topics, allowing to group documents to the more likely matching topics. This organization structure can help identifying which topics are capturing more attention from researchers and also to find gaps for future research. TM can be used indiscriminately, by looking for the most overall referred words, or through the use of specific dictionary words. Since this work is about a focused literature analysis, a dictionary of terms in both banking and BI domains is used.

2.2. Banking

Banks are institutions that operate in the financial business domain, concerning activities such as loaning, deposits management and investments in capital markets, among others. The banking industry is crucial for the economy and thus it is a subject of great interest for researchers in a widespread of different domains, such as management science, marketing, finance and information technologies. [Berger \(2003\)](#) found evidence of a relation between technological progress and productivity in banking. The same author also emphasizes that banks employ statistical models based on their financial data for different purposes, such as credit scoring and risk evaluation.

Financial sector reforms allowed an increase in competition, turning bank lending an important source of funding. Credit risk evaluation is by its own a vast domain, encompassing a large number of research publications within banking and spread through the last twelve years (e.g., [Marqués et al. \(2012\)](#)). Other banking related subject where research has been active is fraud prevention and detection in traditional banking services (e.g., [Abbasi, Albrecht, Vance, & Hansen \(2012\)](#)) and in new communication channels that support e-banking services (e.g., [Shuaibu, Norwawi, Selamat, & Al-Alwani \(2013\)](#)), from which electronic mail spamming in order to illicitly obtain private financial information is a specific case of interest (e.g., [Amayri & Bouguila \(2010\)](#)). E-banking is also subject of another research domain related to technology acceptance regarding new communication channels adopted by banks (e.g., [Vatanasombut, Igbaria, Stylianou, & Rodgers \(2008\)](#), [Lin \(2011\)](#)). A not so recent theme that however has boomed in research, driven by the global financial crisis, is bankruptcy and related subjects such as systemic risk and contagion (e.g., [Hu, Zhao, Hua, & Wong \(2012\)](#)). Competition had also an effect on client related areas, with banks increasing investment in customer retention, customer relationship management (CRM) and targeting (e.g., [Karakostas, Kardaras, & Papatthanassiou \(2005\)](#)).

Research in banking is currently an interesting domain of research. Due to advances in information technology, virtually all banking operations and procedures are automated, generating large amounts of data. Therefore, all the subjects mentioned above can potentially benefit from BI solutions.

2.3. Business intelligence

BI involves several distinct areas and technologies that converge in the common goal of having access to data in order to help businesses by facilitating knowledge and supporting better management decisions. One way to accomplish this is by predicting a certain behavior or result based on data-driven models, in what is known as DM or predictive analytics, thus providing the

Download English Version:

<https://daneshyari.com/en/article/10321908>

Download Persian Version:

<https://daneshyari.com/article/10321908>

[Daneshyari.com](https://daneshyari.com)