Analytical evaluation of emerging scientific trends in business intelligence through the utilisation of burst detection algorithm

Iman Raeesi Vanani and Seyed Mohammad Jafar Jalali*

Department of Industrial Management,
Branch of Information Technology Management,
Allameh Tabataba'i University,
Tehran, Iran
Email: imanraeesi@atu.ac.ir
Email: mohammadjj.it@gmail.com
Email: jalali931@atu.ac.ir
*Corresponding author

Abstract: Business intelligence has become mainstream in recent scientific research trends. The purpose of this research is to study the emerging and fading themes of the business intelligence domain through an analytical overview of keywords, titles and abstracts. Among scientometrics methods for representing the emergent and disappearing trends, the 'burst detection' algorithm has been chosen and applied to the current dataset of high-ranked international papers which can help scholars and practitioners to understand a better overview of business intelligence field by visualising the changes in a recent time period. For this purpose, the data related to business intelligence has been gathered from Web of Science (WoS) core collection dataset between the years 1980–2014 and the burst detection algorithm has been applied on the 'abstract', 'title' and 'keywords' of the dataset which has shown interesting informative results for the future researchers to concentrate on.

Keywords: business intelligence; emergent trends; scientometrics; burst detection.

Biographical notes: Iman Raeesi Vanani is an Assistant Professor of Management Science at the Allameh Tabataba'i University. He received his PhD in IT Management from University of Tehran. His research interests include business intelligence, data mining, and ERP. On these topics, he worked in several national projects. He is also an author of several articles and papers in international journals and conferences about management and business science.

Seyed Mohammad Jafar Jalali is a Research Fellow at Allameh Tabataba'i University in Information Systems. He graduated in Information Technology Engineering and he received his Master of Science in Management of Advanced Information Systems in 2016 at Allameh Tabataba'i University, Iran, under the supervision of Dr. I. Raeesi Vanani. His thesis topic is focused on the

study of the business analytics emerging trends using text mining approaches. His research interest is focused on machine learning, data analysis, business analytics and scientometrics. He has also designed and developed nearly 40 websites over the last few years.

1   Introduction

In information technology age, business intelligence or BI, because of its multi-disciplinary perspectives, has been widely accepted as a recognised attitude for industry and academia. The term business intelligence was coined by Howard Dressner in 1989, However, today it is defined as "a data-driven DSS that combines data gathering, data storage and knowledge management with analysis to provide input to the decision process" (Negash and Gray, 2008). In order to make better sense of BI, it is necessary for a scholar to understand a general overview of this interesting research area, emergent and fading themes and how BI is developing in a whole time-span. For answering these questions, scientometrics can provide a response to these ambiguities. The main purpose of this research is to conduct a scientometrics analysis of 2,124 academic documents in BI field.

Scientometrics is a quantitative approach for evaluating the scientific scholars' activities and is widely used by scholars as a firm basis of research in recent years. The term of scientometrics was firstly used by Nalimov and Mulchenko (1969), however what is called now scientometrics is indebted to Garfield's idea when he introduced for the first time the concept of scientometrics in 1979 (Garfield and Merton, 1979).

It should be noted that drawing visualisation of knowledge, understanding the knowledge domain of a scientific field, using information science as well as mapping of the science during a time period of the area of scientific publications from past, present and future helping to scholars for understanding the science direction and identifying the emergent trends (Börner et al., 2003; Hook and Börner, 2005; Rip, 1988) are the main targets of scientometrics (Rubin, 1998; Cahlik, 2000). This research is intended to analyse and visualise business intelligent emerging areas, using scientometrics techniques, which provides a holistic perspective for academia.

Considering the importance of this issue, there have been limited studies in the past to demonstrate the emergent trends in BI. The major contribution of this study, however, is to investigate the emerging and fading themes of BI, by applying a scientometrics method 'burst detection' to the three key parts of each paper: abstract, title and keywords. Analysis of the burst detection algorithm has been applied using Sci2 tool (Sci2-Team, 2009), a well-known scientometrics software which benefits from open source rules.

Burst detection algorithm is a method for identifying the emergent trends which was introduced first by Kleinberg (2003). The burst of a word means a sudden increase in the frequency of the individual word to detect emergent trends in research publications over a time period, which has been applied to illustrate the rapid and large changes on the abstracts, titles and keywords in this research.

The layout of this research study is organised as follows: Section 2 presents the methodology steps including how to collect data from a well-known scientific database 'Web of Science' (WOS) dataset, the data preparation methods and demonstrating and

discussing some important empirical evaluation results. Section 3 presents the discussion and finally, the study concludes with a summary and some future research directions in Section 4.

2   Research methodology

This section sets out to explain the methodology of this paper which is designed in four sections: data collection, data pre-processing, analysis and discussion.

2.1   Data collection

The data have been retrieved from the database of the Web of Science Core Collection from Thomson Reuters Institute. The term 'business intelligence' has been applied in the search engine of WOS in the category of 'topic' which is the widest category among the other fields, containing title, abstract, year of publication, author keywords and keywords plus from 1980 to 2014.The reason for choosing the time-span of 1980–2014 is the fact to not lose any papers in the field of BI. A total of 2,124 documents were found with the mentioned corresponding search method. The analysis of this paper has been applied on the three important parts of a scientific paper which are 'abstract', 'title' and 'keywords'. Three networks of 'abstract', 'title' and 'keywords' have been constructed and burst detection algorithm has been applied to them.

2.2   Data pre-processing

Main pre-processing methods have been applied on the networks namely lowercasing, tokenising, stemming, and omitting the words which are not helping in concluding the burst analysis outputs. The result is a normalised and cleansed network of abstracts, titles and keywords. The burst detection algorithm has been applied on the pre-processed dataset. The final output of the algorithm is provided in Figures 1, 2 and 3 and Tables 1, 2 and 3. It should be noted that the Tables 1, 2 and 3 have been arranged based on the weights of the burst detection algorithm.

2.3   Analysis

The analysis and discussion part explains the application of the burst detection algorithm to the networks of abstracts, titles and keywords. As the algorithm has been generated in the foresaid networks, a table is created that includes the parameters of word, weight, length, start and end. Word is the burst term that has been extracted from output of burst detection algorithm. The weight parameter indicates the extent of the change in the word frequency that triggered the burst. Tables 1, 2 and 3 have been sorted according to weight. Length is measured via subtracting the end and start parameters.

The visualisation of the tables is shown in Figures 1, 2 and 3. The visualised burst detection diagram is based on two axes. In the X axis, there is 'year' and in Y axis, there is 'emergent subjects'. Finally, the resultant figures have been visualised using temporal bar graphs called horizontal diagrams according to analysis algorithms over a time period.

In the following analysis part, analysis of titles, abstracts and keywords has been discussed.

2.3.1 Analysis of titles

As burst detection algorithm has been applied on the network titles, analysis indicates that it commences in 1958. As an example, the start of burst analysis begins with the term 'business intelligence'.

Table 1 The top burst titles

Word	Weight	Length	Start	End
business intelligence	17.6	39	1958	1996
big	13.91069998	2	2013	2014
mine	9.882628112	7	1998	2004
cloud	9.055472925	2	2013	2014
knowledg	9.047502158	5	2002	2006
system	6.57509322	1	2012	2012
data	5.517525142	4	1998	2001
web	5.432713432	6	2002	2007
social	5.321374647	3	2012	2014
analyt	5.204526505	3	2012	2014

Figure 1 Top visualised burst of titles
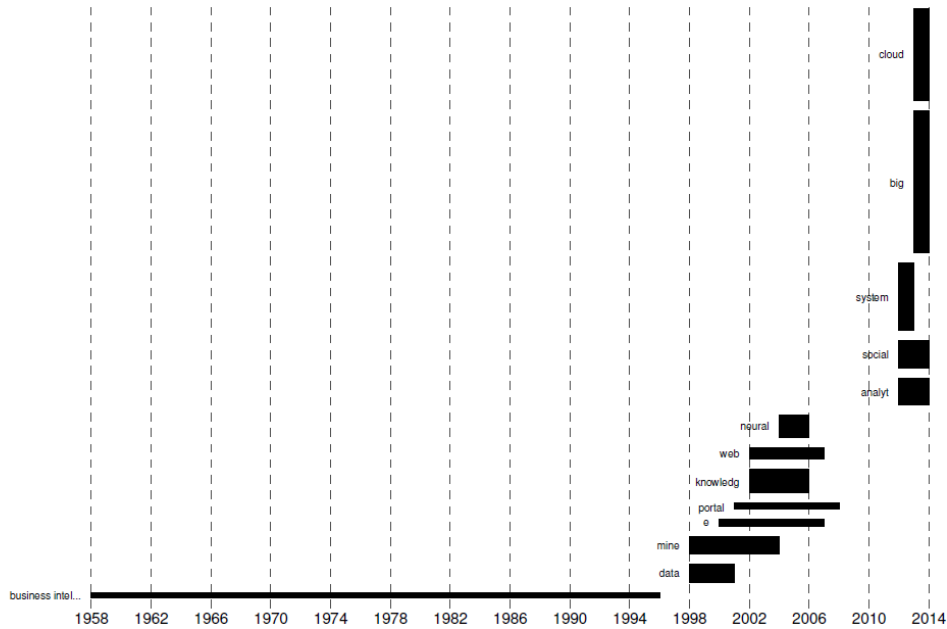
Two burst terms, 'busi' and 'intellie', or 'business' and 'intelligence', were manually merged and changed into 'business intelligence'. To understand this dataset's temporal dynamics, bursts, i.e., sudden increases in the frequency of words in titles, abstracts, and keywords terms were identified. The results are shown in Figure 1; each bursting word is shown as horizontal black bar with a start and ending time, which has been sorted by burst start year. The bar's area represents burst strength. The word 'business intelligence' burst first and begins from 1958 till 1996. Between 1998 and 2008, more words are bursting than in any other period, and the top seven bursting terms ('data', 'mine', 'e', 'portal', 'knowledge', 'web' and 'neural') appear in this time span. Words 'analyt', 'social', 'big' and 'cloud' are still ongoing. The amount of bursting words per year suddenly increased after 2000. Only one bursting word appeared from 1958 to 1996, while 12 bursting terms appeared in the following 15 years. The maximum and minimum weights are devoted to 'business intelligence' and 'portal' respectively, which shows the scholars' focus in BI area according to Table 1.

2.3.2 *Analysis of abstracts*

As it can be seen from Table 2 which is related to analysis abstract of words, contain more term in despite to titles and keywords because of the nature of abstracts, hence the burst words have been seen respectively and could have been done a more in-depth analysis.

Table 2   The top burst abstracts

| Word	Weight	Length	Start	End
big	21.21218816	2	2013	2014
cloud	10.51111588	3	2012	2014
describ	8.916664579	6	2000	2005
general	8.697758174	10	1997	2006
mine	8.64437388	7	1998	2004
commerc	8.556234037	8	2000	2007
social	8.380027645	2	2013	2014
comput	6.906817114	2	2013	2014
portal	6.882144015	6	2001	2006
practition	6.124414686	1	2014	2014

According to Figure 2, the trends have been divided into two parts which are from 1997 to 2008 (the first part) and from 2009 till 2014 (the second part).

In the first, it has been faced with the terms particularly 'internet' which (Roberts, 1999) mentioned that in the scholars' papers as well as the terms 'mine', 'commerce', 'warehouse', 'neural' that have attracted the attention of scholars in that decade (1997–2008). Some words like 'warehouse' have been observed more than twice which reveals the importance of data warehousing which is one of the most significant components of BI.

The second part has been devoted to terms like 'ERP' which refers to enterprise resources planning, and 'cloud' technologies as well as 'big' which has been related to the concept 'big data' (Zikopoulos and Eaton, 2011).

The ongoing terms 'author', 'publish', 'literature' have demonstrated that the scholars of the field of BI have focused on the academic part of BI in these days.

Figure 2 Top visualised burst of abstracts
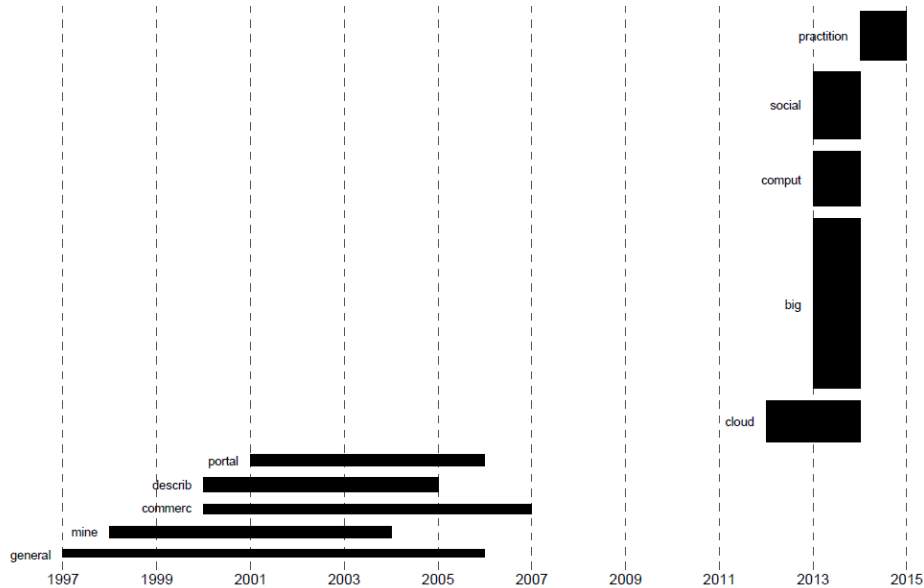
2.3.3 Analysis of keywords

Table 3 demonstrates the most important keywords that have been used by authors in this scientific field as a whole which have included 'e' technologies namely e-commerce, 'mine' data mining and its subfields, 'knowledge' or knowledge management, 'warehouse', 'social' or social networks and finally the word 'big' which refers to 'big data' which has been the focus of many scholars in these recent days.

Table 3 The top burst keywords

Word	Weight	Length	Start	End
big	13.91069998	2	2013	2014
wareh	6.57509322	1	2012	2012
e	6.342548056	17	1991	2007
knowledg	6.339307355	5	2004	2008
mine	5.446372783	7	1998	2004
social	5.309428938	3	2012	2014

Figure 3 Top visualised burst of keywords
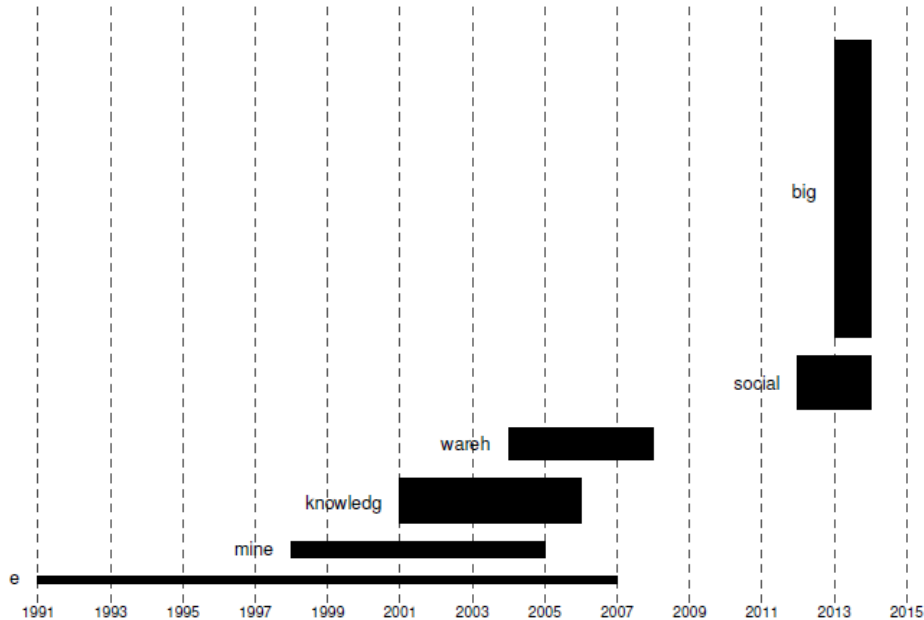
According to the burst analysis provided in Figures 1, 2 and 3, there has been a steady trend in BI publications, for more than four decades (1960s to 1990s), with the exact phrase BI or other similar wordings like business analysis, meta-heuristics, data dredging, data fishing, data analysis, and so on. After that period, a more in-depth approach has been taken into the realm of BI with the emergence of optimised data analysis algorithms with which scholars and researchers could investigate hidden trends in various diversified datasets.

This trend has continued from 1995 to 2006 where the topics around internet-based networks and knowledge management gained momentum. With the emergence of web-enabled databases and massive storage of data through the user involvement in social networks, the trend of publications converged towards a more systematic approach that was knowledge discovery and management over the web.

The advent of social networks gradually led to the emergence of big data. Big data is essentially a huge dataset that can be stored in similar or different algorithms than previous database management systems, however, the major problem is how to deeply analyse and delve into hidden streams and trends that users have left while exploring the web and social networks. In other words, the phrase big data has been coined with another phrase, i.e. data analysis or better to say, business analytics.

Cloud computing has also fuelled the publication trend with huge distributed processing power available to many users around the world. This event has made access to real-time managerial dashboards and analytics faster and easier. The current publication trend illustrates a converging wave of cloud computing, big data, and advanced analytics into a mainstream in which scholars will analyse huge datasets over the web through advanced data mining algorithms.

2.3.4 Brief analysis of words

There are some similarities and differences between the 'title', 'keyword' and 'abstract' analysis. For instance, reviewing the terms used in article abstracts provides a different but complementary illustration, with greater variation between the burst analyses of titles and abstracts shows that the commencement of burst of the words 'portal', 'big', 'social' has happened in the same period of time. Except for the term 'business intelligence' in title analysis, the word 'big' has the most weight among all the terms that have been used in titles and abstracts as well as it shows that the 'big' term which has been referred to 'big data' has the most attention in the scholar community of BI field which begins in 2013.

The results for abstract burst analyses are in contrast to results found for the title burst analysis. The time period for title burst analysis (1958–2014) is compared to abstract burst analysis (1997–2014) in the research. This suggests that the bursts identified in the 1958–2014 period that occurred after 1998, are particularly strong enough to overcome a declining total yearly word count, assuming a relatively consistent average title length over the complete time period. From the most frequent terms used in the abstracts, we can see a broad focus on process, methods and models for BI. Evidence of the primary areas of application for the field can also be seen through the terms, cloud, big, portal, social, and mine. Other frequent terms indicate a focus on technologies, methods, implementations, and applications. As discussed above, the focus of BI research continues to change.

It should be noted that the analysis of abstracts and keywords has examined changes in the use of specific terms, rather than the concepts that these terms may signify. Over time, the ways in which terms are used may change, and different areas of specialisation may use different terms to refer to a similar concept. It has been understood from the results of the figures and tables that while a word has a distinguishable level of seclusion, the possibility of indication of the trend of changes could be more obvious to be extracted and discussed in BI research.

The combination of title and keyword tables' analysis is sensitive to future changes. For instance, words 'social', 'knowledge' and 'big' almost have the same time period with the same weight, which indicates that there is a relationship between the titles and keywords in that specific period while in the future, with the advent of new words and phrases, the instances of words co-occurrence will certainly change.

3 Discussion

From the burst analysis of titles, it can be easily seen that BI has been there for a long time, before it separates into other sub-branch scientific mainstreams. Since the start of the new millennium, in-depth knowledge of occurrences and the causes and reasons behind the events that has taken place in businesses, forced companies to go through more sophisticated algorithms, which were later named data mining, knowledge discovery, meta-heuristics and so on.

After a few years, such sophisticated algorithms highlighted the path of predictive analytics for prosperous business. Predictive algorithms helped businesses acquire new forecasting capabilities beside routines of data gathering and analysis. At the core of

predictive analytics lies the concept of neural networks which is indicated in the visualised burst of titles.

After 2010, there is a major shift in knowledge discovery and utilisation with the emergence of massive social networks which have led to real-time online social interaction and knowledge exchange. The advent of cloud computing and capabilities of running social networks over cloud servers all over the world, has led to the emergence of high volume, high velocity, and high variety data storage, as coined by many authors as big data management.

The integration of big data management, cloud computing, social networks, and predictive analytics is the next mainstream which provides strong infrastructure for BI systems to prove themselves as crucial means of improving intelligence and predictive power of businesses over their competitors.

In the new realm of online social cloud-based BI, the practical usage of BI in social commerce media has gained specific attention and importance. Having good products and services with a high quality, but without a good reach to appropriate customers and prospects means losing many opportunities as well as the prosperity of business. Therefore, BI practices should be well-prepared for new challenges in social communities and social commerce in the next step ahead.

Another aspect of significance lies within the technologies used for improving the level of intelligence. The new technologies should provide a rich, yet simple user experience for business managers all over the business segments. A fast, rich, easy-to-use, and useful BI system will be the most critical component of every new and well-established business to survive in the turbulence of competitive market.

4   Conclusions

In this study, the emerging and fading trends of BI with regard to an analysis of 2,124 articles, retrieved from WOS has been analysed and discussed. The burst detection algorithm has been applied on abstracts, keywords and titles of each document in the BI domain. The visualisation ability of burst detection algorithm has dramatically helped to reach a holistic and comprehensive perception over various aspects of BI concepts and practical terms. Moreover, the main sections of all papers, including abstracts, titles and keywords, have been used for analysis which can provide a complete perspective of the path within which BI will continue to grow and prosper.

As mentioned, the dataset has been retrieved from Web of Science. For future researches, it is suggested to make an integrated comparison on a seamless dataset from all well-known scientific databases such as Scopus, ACM and other famous journal repositories to complement the results of current research.

References

Börner, K., Chen, C. and Boyack, K.W. (2003) 'Visualizing knowledge domains', *Annual Review of Information Science and Technology*, Vol. 37, No. 1, pp.179–255.

Cahlik, T. (2000) 'Comparison of the maps of science', *Scientometrics*, Vol. 49, No. 3, pp.373–387.

Garfield, E. and Merton, R.K. (1979) *Citation Indexing: Its Theory and Application in Science, Technology, and Humanities*, Vol. 8, Wiley, New York.

Hook, P. and Börner, K. (2005) 'Educational knowledge domain visualizations: tools to navigate, understand, and internalize the structure of scholarly knowledge and expertise', *New Directions in Cognitive Information Retrieval*, pp.187–208, Springer, Netherlands.

Kleinberg, J. (2003) 'Bursty and hierarchical structure in streams', *Data Mining and Knowledge Discovery*, Vol. 7, No. 4, pp.373–397.

Nalimov, V.V. and Mulchenko, Z.M. (1969) *Scientometrics. The Study of the Development of Science as an Information Process*, Science, Moskow.

Negash, S. and Gray, P. (2008) *Business Intelligence*, Springer, Berlin Heidelberg.

Rip, A. (1988) 'Mapping of science: possibilities and limitations', in van Raan, A.F.J. (Ed.): *Handbook of Quantitative Studies of Science and Technology*, pp.253–273, Elsevier, North-Holland.

Roberts, P. (1999) 'Scholarly publishing, peer review and the internet', *First Monday*, Vol. 4, No. 4 [online] http://ojs-prod-lib.cc.uic.edu/ojs/index.php/fm/article/view/661/576.

Rubin, R.E. (1998) *Foundations of Library and Information Science*, Neal-Schuman Publishers, Inc, New York.

Sci2-Team (2009) *Science of Science (Sci2) Tool*, Indiana University and SciTech Strategies [online] http://sci2.cns.iu.edu (accessed 3 April 2015).

Zikopoulos, P. and Eaton, C. (2011) *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*, McGraw-Hill Osborne Media, USA.