

Big data - a review in health sciences

Ayça Kurnaz Türkbën *, Software Engineering Department, Bahcesehir University, Istanbul, Turkey

Emre Türkbën, Software Engineering Department, Bahcesehir University, Istanbul, Turkey

Dilek Karahoca, Software Engineering Department, Bahcesehir University, Istanbul, Turkey

Adem Karahoca, Software Engineering Department, Bahcesehir University, Istanbul, Turkey

Suggested Citation:

Türkbën, A., K., Türkbën, E., Karahoca, D., & Karahoca, A. (2016). Big data - a review in health sciences. *Global Journal of Information Technology*. 6(1), 72-79.

Received 13 January, 2016; revised 20 February, 2016; accepted 07 March, 2016.

Selection and peer review under responsibility of Prof. Dr. Adem Karahoca, Bahcesehir University, Turkey

©2016 SciencePark Research, Organization & Counseling. All rights reserved.

Abstract

Technologies are changing very fast and data has an impact on the change of technology and development of world. Data are obtained by social media, the Internet and mobile technologies. For years, academics, researchers and companies utilize some sources and information to analyze them for their studies and jobs. Increasing usage of mobile devices, social networks, electronic records of customers in public and private sectors have led to increase in data. Obtained massive amount of data is called big data. There are a lot of description of big data in the literature, but simply it can be said that; big data is the data which have a massive size and can be obtained from every environment. One of these environment is health environment and it has grown fastly through that huge amount of data exist in this sector like patients' electronic health record. Health sector has a high cost and decision will be taken as soon as possible and correctly in this sector in which timing is critically important. In this manner, the usage of big data in health is important to increase the quality of service, innovative health operations and decrease the cost. In this study, a brief review of literature has done for the use of big data in health sciences for last five years. Big data's content, methods, advantages and difficulties are discussed in this review study.

Keywords: Health science, Big data, Medicine, data mining

1. Introduction

There is an increase in the use of big data in health sector in last few years. One of the reasons is that moving healthcare from a disease-centered towards a patient-centered model in which patients effectively take an interest in their own care and get service for their individual needs by guidance from their medicinal services suppliers [1]. Applying big data analysis to the health sector, the diseases risks can be forecasted before, the strategies to manage the disease can be developed, and the most appropriate treatments can be selected accordance with condition of the patient. Collecting the patients' data sets and handle these aggregated information with big data algorithms can improve reliability of the diagnosing patients. Big data algorithms can offer you a list of possible robust diagnoses based on the information which have been collected from patients and stored as an electronic medical records. Patients can play an active role in their health management by choosing the right one for them. The other benefit of the big data in health sector is that it prevents wasting effort and it makes possible to reach the same information by medical personnel [2]. Accuracy and timeliness of treatment is vital in health sector. Improved health-care systems makes possible to deliver evidence-based care. Also these health staffs may have the ability to carry out certain tasks without the need for a doctor's help. Caregivers should have a high perception of innovation. They should improve therapies and be interested in new productive approaches. Big data allows medical personnel to this authority.

In this meta-analysis study, articles are selected which have the big data, health science and medicine keywords together to the growing effects and trend of this subject. To research the documents relating to big data in health science we used citation analysis in IEEE, Springer and Science Direct databases. The citation analysis permits scientists to inspect how as often as possible a work has been referred to by different writers, giving one measure of the impact of a writer or of a specific article. By citation analysis method citations are counting to decide the academic effect of a study. The impact of a study is assessed by the citing frequency through the specific time period.

Li, Burnham, Lemley and Britton (2010) studied the comparison of four databases' citation analysis capabilities which are Scopus, SciFinder and Google Scholar and they found that Scopus gets better scope of clinical medicine than Web of Science. Google Scholar can link citation information to individual references and all of them can provide comprehensive citation reports [3].

Cui (1999) used citation analysis method to analyse the linked sources medical libraries web pages of 19 of the top U.S. medical schools [4]. The reason of doing this study is there are many web sites related with health care and it should be evaluated the quality of them. 78 most highly cited web sites were found among thousands of cited links.

Moed (2010) studied to develop a new indicator of journal citation impact. By source normalized impact per paper, it quantifies a journal's relevant reference effect, considering qualities of its legitimately characterized subject field, particularly the recurrence at which writers refer to different papers in their reference records, the rate of developing of citation effect, and the degree to which a database utilized for the evaluation covers the field's literature[5]. It was found that basic journals tend to indicate higher citation potentials than applied or clinical journals. This study gives a few helpful recommendations to researchers leading reference examination and the individuals who need help with ordering their own particular reference records. It educates writers of the insight and need and benefit of utilizing numerous hotspots for, and methods for, recognizing references to writers and journals.

Meho and Yang (2007) did a study and it has huge ramifications for subsidizing organizations and in addition editors and distributors of journals who may wish to utilize reference numbers and rankings to recognize subject specialists to submitted compositions and to decide the effect of tasks and articles they financed or distributed. This study has likewise critical ramifications for the more extensive insightful group as researchers receive the strategies and databases depicted or recorded here to distinguish references that may some way or another stay obscure [6].

2. Method

Purposes of the study are listed below:

- It was aimed to make a meta-analysis study about big data in health science between 2010 and 2015 as a brief review.
- It was researched the documents published between the period of 2010-2015 by using Big Data, Health, Medicine keywords in IEEE, Science Direct and Springer database.

To examine reports with respect to Big Data in Health Science, we utilized Citation Analysis. It has been looked reports distributed and ordered in this database that published between the time of 2010-2015 by using Big Data, Health, Medicine keywords in IEEE, Science Direct and Springer index. 40 documents were found in Science Direct, 67 journals from IEEE and 64 articles from Springer.

Documents were grouped and examined by relation between year and number of documents, most frequently used keywords through the years, number of documents through different publishers, number of citation through years, number of documents through different countries and number of documents through different databases.

3. Results

3.1. Relation between year and number of documents

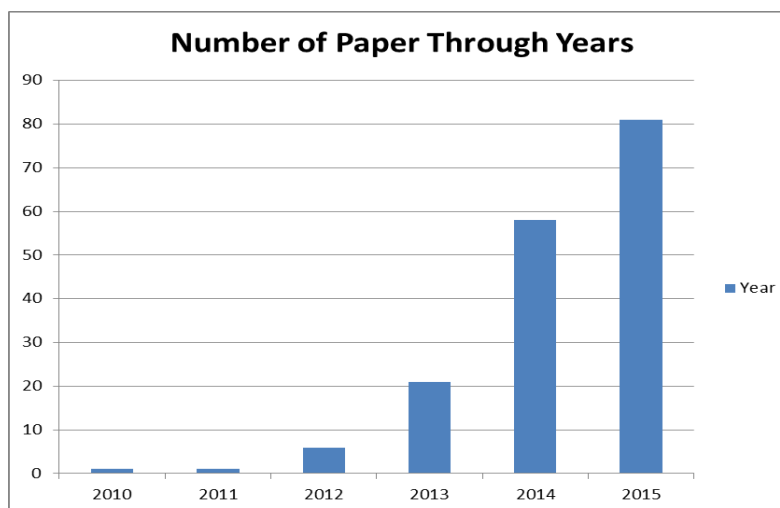


Figure 1. Relation between year and number of documents

The first classification category is number of published paper between 2010 and 2015. In this period, the number of articles has increased over the years and it has reached the highest point in 2015 which is equal to 81 and it equals the almost 50% of the all documents. The increase observed can be interpreted as big data tends to be associated with health science through the years (Figure 1). The minimum number of documents was published in 2010 and 2011.

3.2. Most frequently used keywords through the years

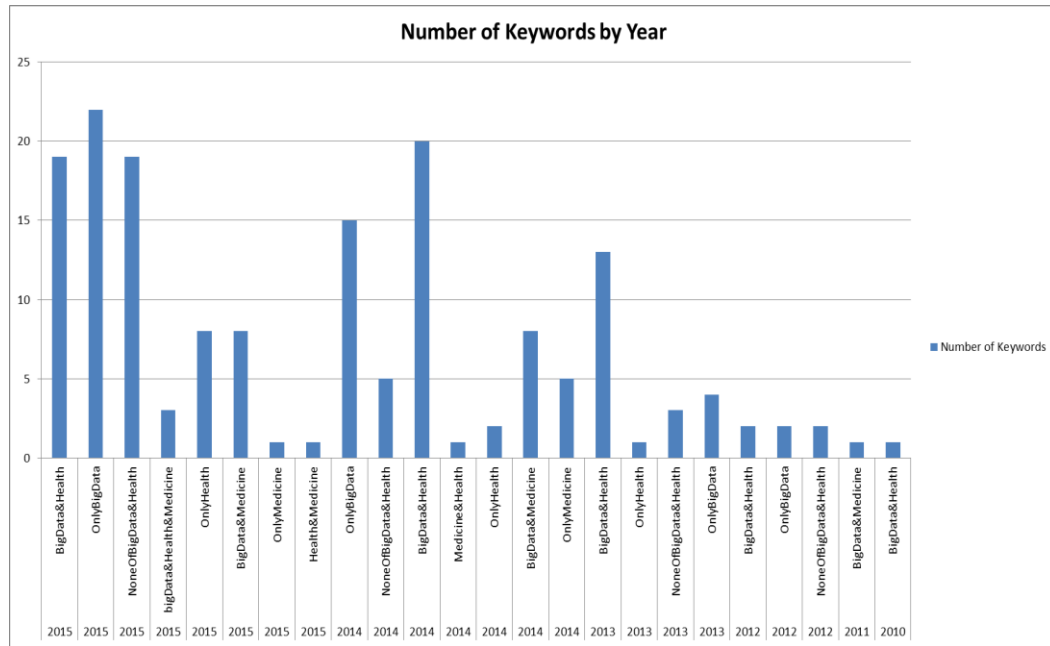


Figure 2. Most frequently used keywords through the years

The most frequently used keywords are presented in Figure 2. In 2015, documents included only big data keyword at most. Big data and health keywords were used almost as much as only big data keywords. Another remarkable point is the number of articles without two keywords is equal to the number of articles that contain both these keywords. Keywords which are big data & health & medicine, only health, big data & medicine and only medicine are not used as often as others in 2015.

In 2014, only big data keyword and big data & health keyword are more preferred than the other keyword combinations. In 2013, the number of articles which includes both big data and health keywords is equal to 62% of documents with other keywords.

3.3. Number of citation through years

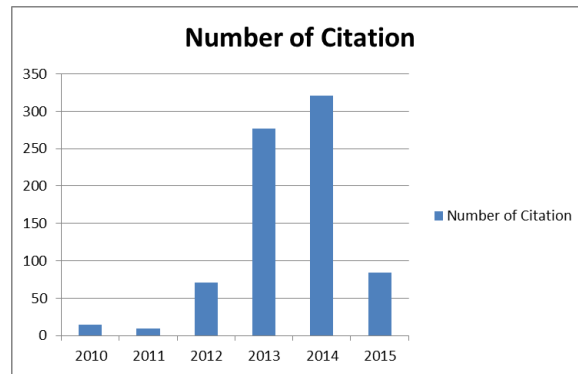


Figure 3. Number of citation through years

The third category is the number of citation through the years. It can be clearly seen that there is a significant increase in 2013 and 2014 (Figure 3). It may be interpreted as researchers started to focus on the articles which combine big data and health science. If it is mentioned the decrease in 2015, that means this study was carried out in 2015, the number is expected to increase next years.

3.4. Number of documents through different countries

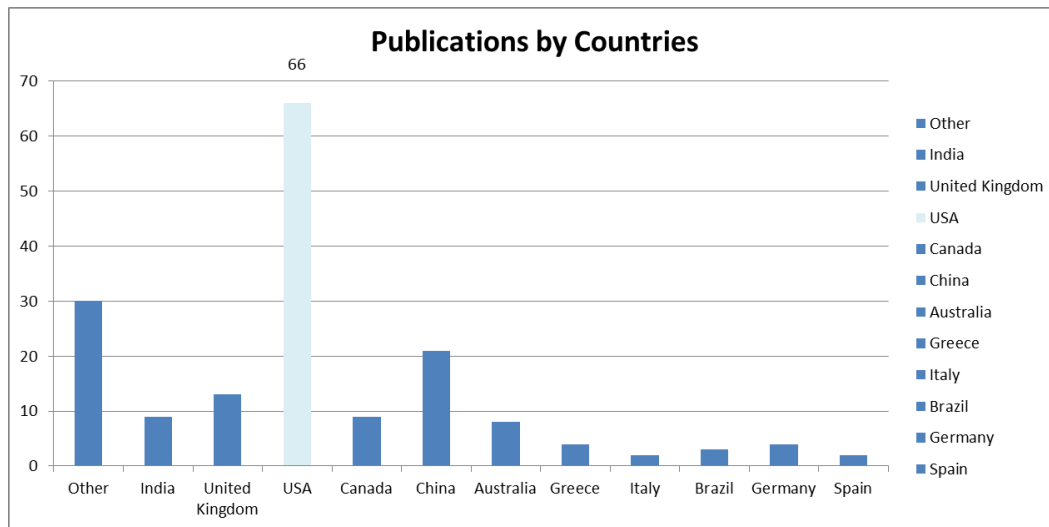


Figure 4. Number of documents through different countries

According to the number of publications by countries figure, the most articles are published by USA. China, UK, Canada, India and Australia come after USA respectively (Figure 4). It is common that these countries published articles together.

3.5. Number of documents through different databases

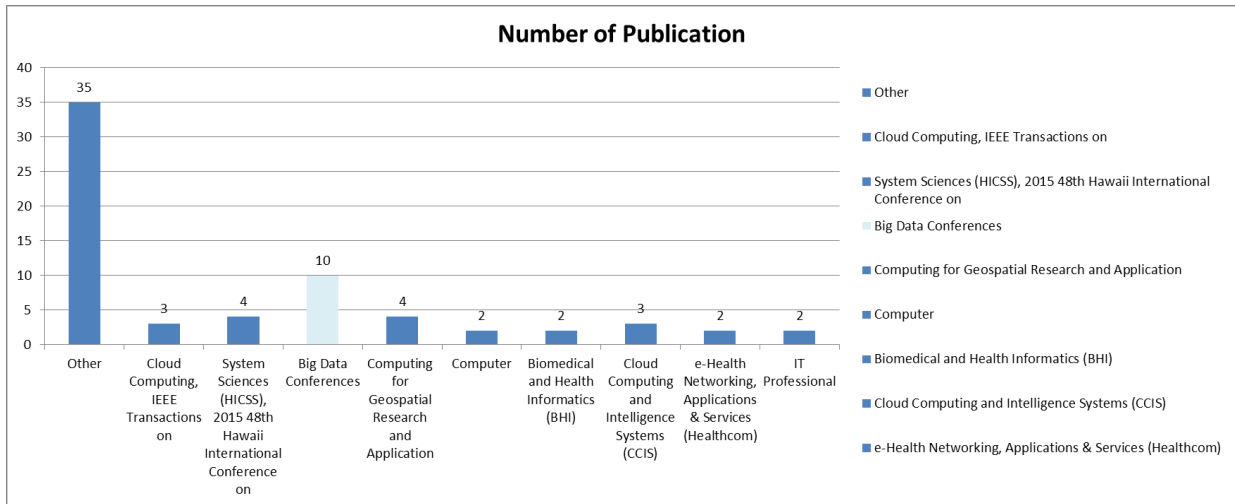


Figure 5. Number of publication in IEEE database

In IEEE databases, the most number of articles are published in Big Data Conferences nearly 52% of all journals and conference publications (Figure 5).

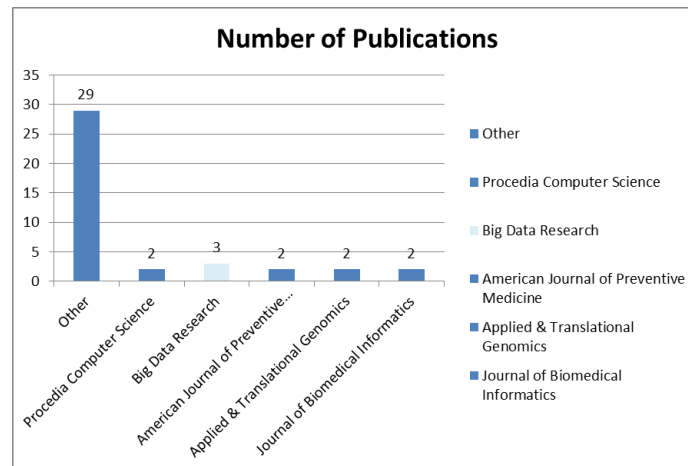


Figure 6. Number of publication in Science Direct database

In Science Direct database, all journals contain approximately the same number of articles with keywords big data, health science and medicine (Figure 6).

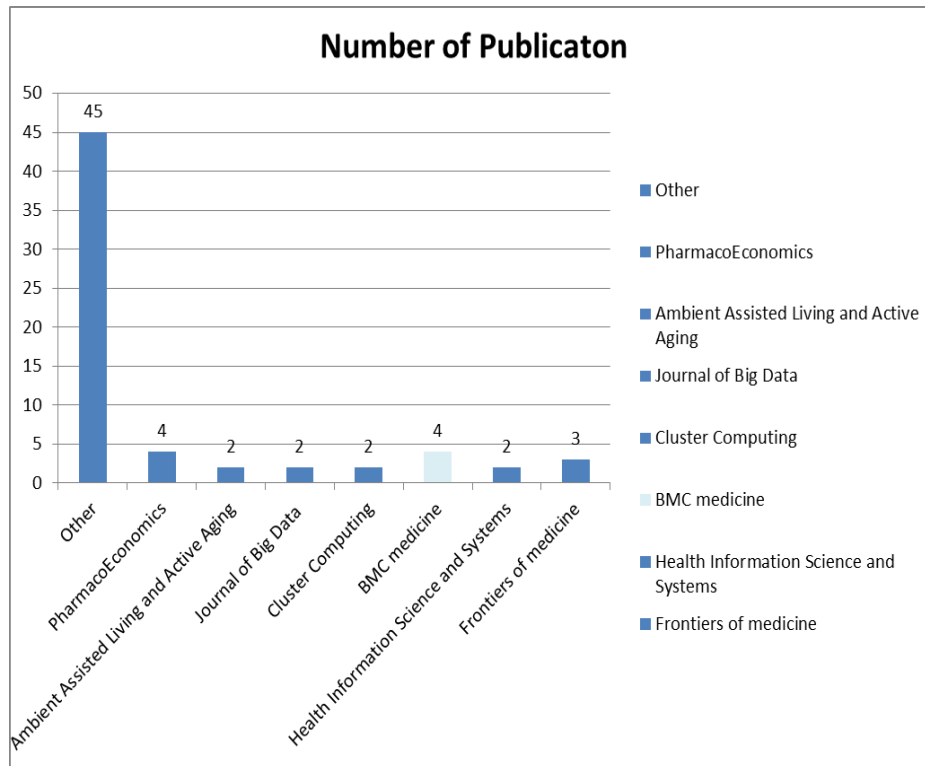


Figure 7. Number of publication in Springer database

Despite PharmacoEconomics and BMC (BioMed Central) medicine journals, related with big data, health science and medicine concepts, contain more articles from other journals with very small margin, all journals include approximately the same number of articles in Springer Databases (Figure 7).

4. Conclusion

As a result of this meta-analysis study, recent years a significant increase was observed in publications about big data in health science, especially between 2013 and 2014. There are a lot of different types of journals which includes documents about big data in health science. A lot of authors and researcher from different countries are working about this topic. Especially, USA makes a significant difference about number of publications in this topic. It was observed, there is an increase in the citation frequency towards 2015. Different databases contain numerous articles in this topic and IEEE database contains significantly more publications about big data and health. We can say that, the publications about our topic will continue to improve. Different models, analysis and systems will be developed through the changes of technology in health sector. Also the increase in the number of studies about health sector related big data, will help this sector to innovate some new technologies.

References

- [1] CHAWLA, Nitesh V.; DAVIS, Darcy A. Bringing big data to personalized healthcare: a patient-centered framework. *Journal of general internal medicine*, 2013, 28.3: 660-665.

- [2] KAYYALI, Basel; KNOTT, David; VAN KUIKEN, Steve. The big-data revolution in US health care: Accelerating value and innovation. *Mc Kinsey & Company*, 2013.
- [3] Li, J., Burnham, J. F., Lemley, T., & Britton, R. M. (2010). Citation analysis: Comparison of web of science®, scopus™, SciFinder®, and google scholar. *Journal of electronic resources in medical libraries*, 7(3), 196-217.
- [4] Cui, L. (1999). Rating health Web sites using the principles of citation analysis: A bibliometric approach. *Journal of medical Internet research*, 1(1).
- [5] Moed, H. F. (2010). Measuring contextual citation impact of scientific journals. *Journal of Informetrics*, 4(3), 265-277.
- [6] Meho, L. I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *Journal of the american society for information science and technology*, 58(13), 2105-2125.