Powerful things to mix: Big Data, Text Mining, Social Networks Analysis (SNA) and Biomedical Research.

Caio Fernandes Moreno
Master in Artificial Intelligence
Universidad Politécnica de Madrid
Boadilla del Monte, 28660 Madrid (Spain)
caio.fmoreno@alumnos.upm.es
caiomsouza@gmail.com

ABSTRACT

During many years companies and universities have been investing in biomedical research in different countries, but it is still hard for individual or groups of investigators to collaborate in a very effective way.

The internet created a world of opportunities and connected people everywhere. Social networks analysis (SNA) help us understand patterns of interaction between social entities. A number of SNA studies have shed light on the characteristics of research collaboration networks (RCNs). Especially, in the Clinical Translational Science Award (CTSA) community, SNA provides us a set of effective tools to quantitatively assess research collaborations and the impact of CTSA. [1]

Although the huge increase in the number of collaborations between the scientific community it is still extremely difficult to use all this produced biomedical knowledge and content. One efective way to approach this Big Data problem is to apply text mining and big data to help researchers accelerate identification of for example: novel drug candidates and novel drug targets. [5]

Author Keywords
Social networks analysis, biomedical research, RCN, CTSA, SNA, Text Analytics, Big Data.

1.- INTRODUCTION

Researchers are under pressure all over the world to innovate faster than ever. Big data, Text Mining and Social Networks Analysis (SNA) offer the promise of unlocking novel insights and accelerating breakthroughs in the Biomedical field. Ironically, although more data are available than ever, only a fraction is being integrated, understood, and analysed.

Cognitive solutions can understand different types of data such as lab values in a structured database or the text of a scientific publication. Cognitive solutions are trained to understand technical, industry-specific content and use advanced reasoning, predictive modelling, and machine learning techniques to advance research faster.

Watson, a cognitive computing technology, has been configured to support life sciences research. This version of Watson includes medical literature, patents, genomics, and chemical and pharmacological data that researchers would typically use in their work. Watson has also been developed with specific comprehension of scientific terminology so it can make novel connections in millions of pages of text. Watson has been applied to a few pilot studies in the areas of drug target identification and drug repurposing. The pilot results suggest that Watson can accelerate identification of novel drug candidates and novel drug targets by harnessing the potential of big data. [5]

There is a huge potential in the present and in the future for integrating Big Data, Text Mining, Social Networks Analysis (SNA), PubMed and all different kids of sources to save life.

2.-BASIC CONCEPTS

Biomedical Research is the area of science devoted to the study of the processes of life, the prevention and treatment of disease, and the genetic and environmental factors related to disease and health.

PubMed comprises more than 26 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites.[6]

Big data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation. [7]

Social network analysis (**SNA**) tools are used to analyze patterns of relationships among people in groups. They are useful for examining the social structure and interdependencies (or work patterns) of individuals or organizations. SNA involves collecting data from multiple sources (such as surveys, emails, blogs and other electronic artifacts), analyzing the data to identify relationships, and mining it for new information such as the quality or effectiveness of a relationship. Organizational network analysis is the form of SNA that examines the information flow among individuals. It depicts the informal social network typically of groups working in the same enterprise. Value network analysis (VNA) examines the deliverables exchanged among roles typically groups of people from multiple organizations that need to work together. Social influence network analysis scans social media to identify influential people, associations or trends in the collective. [8]

Text Mining is the process of extracting information from collections of textual data and utilizing it for business objectives. [9]

Artificial intelligence is technology that appears to emulate human performance typically by learning, coming to its own conclusions, appearing to understand complex content, engaging in natural dialogs with people, enhancing human cognitive performance (also known as cognitive computing) or replacing people on execution of nonroutine tasks. Applications include autonomous vehicles, automatic speech recognition and generation and detecting novel concepts and abstractions (useful for detecting potential new risks and aiding humans quickly understand very large bodies of ever changing information). [10]

3.- STUDY CASES

A.- CollaborationViz

As we can see in the article by Jiang Bian et al. [1] they developed a tool called CollaborationViz to help nontechnical biomedical researchers and administration to understand the nature and evolution of collaboration.

CollaborationViz has been created to not only better disseminate the results of the authors network analyses on biomedical RCNs, but also to support visual analytics. A live demo of CollaborationViz can be found at http://bianjiang.github.io/rcna/. [1]

Figure 1: The main interface of CollaborationViz, an interactive visual analytical tool for exploration of biomedical research collaboration networks.

Clinical translational science embraces inter-disciplinary collaborations. [1]

One of the key objetives of the Clinical Translational Science Award (CTSA) is to promote cross-disciplinary collaborations that can accelerate the translation and application of biomedical research discoveries into clinical settings. [1]

When we think about a research collaboration community or research collaboration networks (RCNs), we can also understand them as Social Networks with social entities. Seeing them this way, helps us understand the possibility of applying Social Network Analysis (SNA) techniques to understand relationships (e.g co-authorships in scientific publication and collaborations on grants).

Social network analysis (SNA) has been deemed as an effective tool to assess inter- and intra-institution research collaborations in the CTSA community [1 -2]

B.- IBM Research Accelerated Drug Discovery [4]

Recent estimates suggest that the costs of bringing a new drug to market has reached $2.5 billion and >12 years of investment.3 Of drug candidates, 80% to 90% fail to gain U.S. Food and Drug Administration approval.4 The most common reasons for failure include lack of efficacy, lack of safety, poor dosage selection, and poor endpoint selection.4 and 5 Looking across disease states, in some therapeutic areas approval rates are as low as 6.7%.6 [5]

There are nearly 200,000 active clinical trials, 21,000 drug components, 1357 unique drugs, 22,000 genes, and hundreds of thousands of proteins. [11 and 12]

Each of these areas of study includes testing and experiments that yield vast quantities of data, making it difficult for any 1 researcher or even teams of scientists to absorb.[12, 13 and 14] There are >24 million published medical and scientific articles in the 5600 journals in MEDLINE alone, with 1.8 million new articles published annually. [15 and 16] Meanwhile, the average researcher typically reads on average 250 to 300 articles in a given year. [17] This suggests that scientists may not be keeping up with the basic science published in their area of specialty, let alone making novel connections that could come from harnessing many data sources. The volume of published science grows at a rate of ~9% annually, doubling the volume of science output nearly every 9 years. The ability to absorb only a fraction of available information results in many lost opportunities to further research. Drug discovery depends on identifying novel and effective targeting strategies that produce better clinical outcomes for patients. Harnessing volumes of information about how disease processes originate and progress and how drugs affect animals and humans could yield novel treatment strategies. [5]

Y. Chen et al[5] explained in her article the IBM research project lead by her where they used technology to accelerate the drug discovery process.

5.- CONCLUSION

We are entering the Cognitive Era as promoted by IBM. Every time more we will use technology to approach big challenges in our life.

Artificial Intelligence is shaping our world and the way we live and we will live. Machines are becoming smarter. In every science field we are seeing progress, we live in a world where we can buy today a Self-drive Tesla Car that are already allowed by the U.S government to drive alone in a Highway. [3]

The future is brilliant, every where we will increase our power to analyse data and make the world a better place.

Researchers can already use technologic tools to help them increase productive in research projects as we could see in theses two examples.

There is a really powerful mix between Big Data, Text Mining, Social Networks Analysis (SNA) and Biomedical Research.

The world is becoming a better place and things are going faster and faster.

REFERENCES

1. Bian J, Xie M, Hudson TJ, Eswaran H, Brochhausen M, et al. (2014) CollaborationViz: Interactive Visual Exploration of Biomedical Research Collaboration Networks. PLoS ONE 9(11): e111928. doi:10.1371/journal.pone.0111928

2. CTSA Consortium: Evaluation Key Function Committee (2011) Evaluation - social network analysis. Available: https://www.ctsacentral.org/committee/evaluation-social-network-analysis. Accessed 2014 April 18.

3. Davos 2016 - The State of Artificial Intelligence. Available: https://www.youtube.com/watch?v=VBceREwF7SA. Accessed 2016 May 28.

4. IBM Research Accelerated Drug Discovery. Available: https://www.youtube.com/watch?v=lTB8BTuteIs. Accessed 2016 May 28.

5. Chen Y, Elenee Argentinis and Griff Weber. (2016). IBM Watson: How Cognitive Computing Can Be Applied to Big Data Challenges in Life Sciences Research. Clinical Therapeutics. Volume 38, Issue 4, April 2016, Pages 688-701. doi:10.1016/j.clinthera.2015.12.001

6. PubMed. Available: http://www.ncbi.nlm.nih.gov/pubmed. Accessed 2016 May 28.

7. Big Data definition. Gartner. Available: http://www.gartner.com/it-glossary/big-data/. Accessed 2016 May 28.

8. Social Network Analysis (SNA) definition. Gartner. Available: http://www.gartner.com/it-glossary/social-network-analysis-sna/. Accessed 2016 May 28.

9. Text Mining definition. Gartner. Available: http://www.gartner.com/it-glossary/text-mining/. Accessed 2016 May 28.

10. Artificial Intelligence definition. Gartner. Available: http://www.gartner.com/it-glossary/artificial-intelligence/. Accessed 2016 May 28.

11. John P. Overington, B. Al-Lazikani, A. Hopkins. How Many Drug Targets Are There? Nat Rev Drug Discov, 5 (2006), pp. 993 996

12. Wu, Po-Yen, Venugopalan, Janani, Kothari, Sonal, et al. April, 2014. Big data analytics biomedical and health informatics for personalized cardiovascular disease care. Lifesciences.ieee.org. http://lifesciences.ieee.org/publications/newsletter/april-2014/539-big-data-analytics-biomedical-and-health-informatics-for-personalized-cardiovascular-disease-care. Accessed May 26, 2015.

13. Philip Chen, C.L. Zhang, Chun-Yang. Data-intensive applications, challenges, techniques and technologies; a survey on big data. Inf Sci, 275 (2014), pp. 314 347

14. National Institute of Health. Trends, charts and maps. https://clinicaltrials.gov/ct2/resources/trends. Accessed August 9, 2015.

15. NIH U.S. National Library of Medicine MEDLINE® Fact Sheet. https://www.nlm.nih.gov/pubs/factsheets/medline.html. Accessed August 9, 2015

16. Eveleth R. Academics write papers arguing over how many people read (and cite) their papers. Smithsonian.com. 25, 2014. http://www.smithsonianmag.com/smart-news/half-academic-studies-are-never-read-more-three-people-180950222/?no-ist. Accessed May 26, 2015.

17. Van Noorden R. Scientists may be reaching a peak in reading habits. February 3, 2014. http://www.nature.com/news/scientists-may-be-reaching-a-peak-in-reading-habits-1.14658. Accessed May 26, 2015.