Recent approaches to creating and analyzing big data in technology management and commercialization

DR. CLOVIA HAMILTON, ASST PROFESSOR OF MANAGEMENT
COLLEGE OF BUSINESS ADMINISTRATION
WINTHROP UNIVERSITY

Background

This is a study of three (3) technological advances related to the use of Big Data: web scraping, natural language processing and machine learning; and ways to use these advances in technological innovation management – and in particular tech commercialization. This research envisions research universities, federal labs and corporations sharing new ideas about technological innovations in one clearinghouse which would be a Big data set and idea management platform.

Web scraping
Use of a bot or web crawler to copy data from the web into a spreadsheet or database for later retrieval and analysis (Vargiu, 2013)

Natural language processing (NLP)
NLP is a branch of artificial intelligence that involves programming computers to process and analyze large amounts of natural, human language data. NLP can be sued to draw insights that can be used in business decision making (Mills, 2018).

Machine learning
This branch of artificial intelligence uses statistical techniques to help users learn from data with use of task specific algorithms. Deep learning makes use of learning data representations rather than the task specific algorithms (Marr 2018).

Findings

Big Data is the use of data sets that require large storage space and time to analyze (Kaisler, 2013; Ward, 2013). The primary beneficial business value-added to big data driven, strategic decision making in open innovation and related strategic alliances in tech development is cost savings, efficiencies and mitigation of risks (World Econ Forum, 2012; Manyika, 2011; Harbinson, 1998; Uddin 2011; Robinson, 2017).

The North Carolina State University (NCSU) Center for Innovation Management Studies (CMIS), with help from IBM, developed a powerful, low cost Unstructured Text Analysis (UTA) platform to help companies make better, more informed decision in identifying data-drive trends in marketing, customer and financial intelligence; looking for root causes of situations; extracting information; enhancing predictive models and identifying key influences (Mugge). Similarly, Spigit developed new machine learning technology to analyze unstructured data across ideas to consolidate similar ideas and to connect people interested in like-minded ideas. The goals is to achieve efficiency gains and to focus teams on advancing the best ideas (Spigit, 2018).

NEC reported investing in data science to create new value from big data using AI for visualization and analysis. They promote open innovation of core technologies in safety, retail and global carrier. They created a Chief Technology Officer (CTO) position to manage their initiatives and an AI Collaborative Research Lab (NEC, 2017).

Why use Big Data in technology innovation and commercialization management?

Researching the patent landscape can help R&D teams decide which opportunities to pursue and which to avoid (Morris). Creating communities for the sharing of new ideas will require a big data privacy defensive strategy. There is a lack of a comprehensive overview for the use of Big data in open innovation strategies (Del Vecchio, 2017).

Big data can be used to profile market influencers and market decision makers (Robinson, 2017). Tools such as Cassandra, Spark, MapReduce and Hadoop are being used.

Research Questions

What will it take? What are the challenges?

Method

This research is exploratory. The chosen method of research is a literature review.

Recommendations

Currently, the average size of a university technology transfer office (TTO) staff is only three (3) persons. The staff does not have expertise in Big Data web scraping, natural language processing or machine learning. The typical American TTO currently lacks the expertise to manage or make use of the proposed Big Data platform. However, large corporations such as GE and Google can (Wenning, 2015).

A clearinghouse Big Data cloud platform should be developed for promoting collaborations among university, federal lab and corporate technological innovation players. This will require that:

- All participants have Chief Technology Officers that are competent in web scraping, NLP and machine learning.
- Attribution and acknowledgement of the first to invent
- Trust among participants
- Standards developed by a new Standards Board

References

Del Vecchio, Pasquale; Di Minin, Alberto; Messeni Petruzzelli, Antonio; Panniello, Umberto; Pirri, Salvatore. (2017). Big data for open innovation in SMEs and large corporations: Trends, opportunities, and challenges. *Creative Innovation Management*, 27, 6-22.
Georgetown University. (2018). Tools for Managing Technology with Big Data. Retrieved 11/2, 2018, from https://scsonline.georgetown.edu/programs/masters-technology-management/resources/tools-managing-technology-big-data
Grainger, Michelle.). CIMS Team Explores Using Big Data for Open Innovation. Retrieved 11/2, 2018, from https://cims.ncsu.edu/cims-team-explores-using-big-data-for-open-innovation/
Harbinson, J.R. (1998). *Smart alliances: A practical guide to repeatable success*. San Francisco, CA: Jossey-Bass.
Kaisler, S.; Armour, F.; Espinosa, J.A.; Money, W. (2013). *Big data: Issues and challenges moving forward*. Paper presented at the 46th Hawaii International Conference on System Sciences (HICSS), Hawaii.
Kayser, Victoria; Nehrke, Bastian; Zubovic, Damir. (2018). Data science as an Innovation Challenge: From Big Data to Value Proposition. *Technology Innovation Management Review*, *8*(3), 16-25.
Long, Mary C. (2017). Where Innovation Management and Big Data Meet. *CMS Wire*.
Ma, Yanhong. (2017, 2018). *Research on Technology Innovation Management in Big Data Environment*. Paper presented at the ICAESEE.
Manyika, James; Chui, Michael; Brown, Brad; Bughin, Jacques; Dobbs, Richard; Roxburgh, Charles; Hung Byers, Angela. (2011). Big data: The next frontier for innovation, competition, and productivity: McKinsey Global Institute.
Marr, Bernard. (2016). *Big Data in practice: How 45 successful companies used Big Data analytics to deliver extraordinary results*. Chichester, UK: Wiley.
Marr, Bernard. (2018). What is Deep Learning AI? A simple guide with 8 practical examples. *Forbes*.
Mills, Terence. (2018). What is Natural Language Processing and what is it used for? *Forbes*.
Morris, Chad.). Trying out a tool that makes patent searches a snap. Retrieved 11/2, 2018, from https://cims.ncsu.edu/trying-out-a-tool-that-makes-patent-searches-a-snap/
Mugge, Paul; Kori, Richard.). Why Big Data is not all Hype: The Power of Unstructured Text Analytics. Retrieved 11/2, 2018, from https://cims.ncsu.edu/why-big-data-is-not-all-hype-the-power-of-unstructured-text-analytics/
NEC. (2017). NEC CSR Report Innovation Management.
Robinson, Robin. (2017). The Big Data Influence on Commercialization. *PharmaVoice*.
Rothe, Rasmus. Bringing machine learning research to product commercialization - Lessons learned at Merantix: The difference between academia and industry. *Medium*.
Spigit. (2018). Spigit unveils the future of innovation management software to help business leaders foster a culture of engagement and innovation. Retrieved 11/2, 2018
Thomas, Llewellyn; Leiponen, Aija. (2016). Big Data Commercialization. *IEEE Engineering Management Review*, *44*(2), 74-90.
Uddin, M.; Lecturer, B. (2011). Strategic alliance and competitiveness: Theoretical framework. *Journal of Arts Science & Commerce*, 43-54.
Vargiu, Eloisa; Urru, Mirko. (2012). Exploiting web scraping in a collaborative filtering-based approach to web advertising. *Artificial Intelligence Research*, *2*(1).
Ward, J.S.; Barker, A. (2013). Undefined by data: A survey of big data definitions.
Wenning, Maria. (2015). Why Big Data Matters in Innovation Management. Retrieved 11/2, 2018, from https://blog.hypeinnovation.com/why-big-data-matters-in-innovation-management
World Economic Forum. (2012). Big Data, Big Impact: New Possibilities for International Development. Switzerland: The World Economic Forum.

Challenges:

- When universities, federal labs and corporations partner, each define their priorities differently (Georgetown University).
- The focus needs to be on the application of Big Data search results (Grainger, 2016); i.e. what would each strategic partner intend to do with the data? What should they each be allowed to access?
- How to analyze decisions made using the Big dataset (Kayser, 2018).
- There needs to be policies in place to address privacy, security, intellectual property rights and liability (McKinsey; Long, 2017)
- Technical management in a big data environment will require extremely high professional, quality staff to ensure good use of the platform and the accuracy of data analysis (Ma, 2017). The current typology for big data management include data suppliers, managers, custodians, aggregators, app developers, and service providers in a disaggregated fashion rather than an ecosystem (Thomas,).
- Getting sufficient access to data, optimizing product market fit, dealing with laws and regulations, and successfully going to market are challenges (Rothe)
- Academic and industry work flows differ. Academics begin with a fixed training dataset such as MNIST or ImageNet to train a specific model on; and to develop novel models or adapt existing methods to improve model performance. Industry use of machine learning begins with a fixed performance requirement (Rothe).