

Big Data Applications in Media and Entertainment Industry

Jiaan Wang
Indiana University Bloomington
3209 E 10th St
Bloomington, Indiana 47408
jervwang@indiana.edu

ABSTRACT

The growth of big data and its various applications in media and entertainment industry has been swift in recent years as well as the rapid surge of big data and the increasing need for big data technologies. We describe the problems that come with big data and its challenges in the industry. We then present various utilization of big data and why big data is important to the advancement of media and entertainment industry.

KEYWORDS

i523, hid233, Big data, Media, Entertainment industry, Technology, Recommendation

1 INTRODUCTION

The amount of data being generated is increasing exponentially every year. Currently, we don't have the resources to process or analyze all the data. For example, giant tech companies like Google process over 20 petabytes of data daily [6]. "The rate at which we are generating data is rapidly outpacing our ability to analyze it and the trick here is to turn these massive data streams from a liability into a strength" [2]. Despite that, the technologies used to collect, analyze and interpret data are continuously improving [6].

IDC, the International Data Corporation, believes that "organizations that are best able to make real-time business decisions using big data streams will thrive, while those that are unable to embrace and make use of this shift will increasingly find themselves at a competitive disadvantage in the market and face potential failure. This will be particularly true in industries experiencing high rates of business change and aggressive consolidation" [7].

But what is big data? Wanda Group, a multinational conglomerate company based in China, defines big data as a DIKW hierarchical model, which stands for Data, Information, Knowledge and Wisdom [9]. "Big data is about the growing challenge that organizations face as they deal with large and fast-growing sources of data or information that also present a complex range of analysis and use problems. Big data technologies describe a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and analysis [7].

Emerging sources for big data include industries that are preparing to digitize their content. Particularly, "the media and entertainment industry moved to digital recording, production, and delivery in the past five years and is now collecting large amounts of rich content and user viewing behaviors" [7].

2 CHALLENGES IN MEDIA AND ENTERTAINMENT INDUSTRY

"The problem with the massive data collection and distribution system created is: big data is a big mess. Most of the data captured in our daily lives just sits around, cluttering up storage space on devices and slowing down connections" [6].

"Media and entertainment industry has frequently been at the forefront of adopting new technologies. The key business problems that are driving media companies to look at big data capabilities are the need to reduce the costs of operating in an increasingly competitive landscape and at the same time, the need to generate revenue from delivering content and data through diverse platforms and products" [4].

Traditional TV media are facing challenges as its data is scattered. "It has internal data from set-top boxes, network management systems, BOSS systems, etc. as well as external data from online user behaviors. Data integration is the primary challenge in big data applications of traditional TV media" [9]. In China, the overall economy of traditional TV media does not look promising. The amount of time user spent on traditional TV has declined while more time is spent on Internet TV. "Studies have shown that, in 2012, Internet TV user base has reached 26.1 million while traditional TV user base is only 600 million. In addition, traditional TV operation rate has decreased from 70 percent in 2009 to 30 percent in 2012" [9].

These are the main challenges media and entertainment industry needs to deal with in order to better utilize big data to make a difference:

- "Making sense of data streams, whether text, image, video, sensors, and so on. Sophisticated products and services can be developed by extracting value from heterogeneous sources" [4].
- "Exploiting big data step changes in the ability to ingest and process raw data, so as to minimize risks in bringing new data-driven offerings to market" [4].
- "Curating quality information out of vast data streams, using algorithmic scalable approaches and blending them with human knowledge through curation platforms" [4].
- "Accelerating business adoption of big data. Consumer awareness is growing and technical improvements continue to reduce the cost of storage and analytics tools among other things. Therefore, it is more important than ever that businesses have confidence that they understand what they want from big data and that the non-technical aspects such as human resources and regulation are in place" [4].

3 APPLICATIONS IN MEDIA AND ENTERTAINMENT INDUSTRY

“Massive quantities of data are already being captured about entertainment. For example, *Supernatural*, an American horror series, created by Eric Kripke in 2005. Now in its seventh season, it has generated roughly 112 hours of footage. We have a lot of pixels and we also have every action of every character; every line of dialogue; a history of when, where, and how often everyone dies. Because all of that information is data, what we actually have, in and around those 112 hours of pixels, is a map to the world of *Supernatural*, and the characters inside it. Today, all of that footage and all of that information is locked away in old style data collections: fixed and unwieldy. But if we can store all that information in a system, modeled more on biology than books, and apply our significant and increasing processing power to analyze and respond to the world, rather than just move it around mechanically, then we have the possibility of generating and interacting with the world and the characters of *Supernatural*” [6].

“Hollywood also uses big data big time” [5]. “IBM worked with a media company and ran its predictive models on social media buzz for the movie *Ram Leela*. According to the reports, IBM predicted a 73 percent success for the movie based on right selection of cities. Such rich analysis of social media data was conducted for *Barfi* and *Ek Tha Tiger*. All these movies had a runaway success at the box office. Shah Rukh Khan’s *Chennai Express*, one of the biggest box office grosses in 2013, used big data and analytic solutions to drive social media and digital marketing campaigns. IT services company *Persistent Systems* helped *Chennai Express* team with the right strategic inputs. *Chennai Express* related tweets generated over 1 billion cumulative impressions and the total number of tweets across all hash tags was more than 750 thousand over the 90-day campaign period. Singapore based big data analytic firm *Crayon* has worked with leading Hindi film industry producers to understand the kind of music to release in order to create the right buzz for movies. In addition, Lady Gaga and her team browse through listening preferences and sequences to optimize the play list for maximum impact at live events” [3].

“Sports is another area where big data is making big impacts. Germany, FIFA 2014 champion, has been using SAP’s *Match Insights* software to analyze team performance which made a big difference for the team. It analyzes data such as player positions, touch maps, passing abilities, ball retention and even metrics such as aggressive play. In addition, Kolkata Knight Riders, an Indian Premier League team, used *Match Insights* to determine the consistency of its players which helped in auction as well as in ongoing training” [3].

“By using big data to understand why the customers subscribe and unsubscribe, entertainment organizations could develop the best product and promotional strategies to attract and retain clients. Unstructured sources best handled by big data apps like email, call detail records and social media sentiment reveal factors that are often overlooked for driving customer interest. Big data makes possible the understanding of consumption of digital media and entertainment and behavior that could be used together with traditional data demographic for personalized advertising in the right context at the right time, in the right place” [5].

“Recommendation engines are very powerful personalization tools because it’s a great way to show people items they will like. A lot of Amazon’s fantastic revenue growth has been built on successfully integrating recommendations across the buying experience from product discovery to checkout” [1].

“Amazon is investing a large amount of talent and resources on getting better artificial intelligence, specifically deep learning technology to make recommendation engines which learn and scale even more efficiently. In May 2016, Amazon opened up its sophisticated artificial intelligence technology as a cloud platform. The company unveiled DSSTNE, an open source artificial intelligence framework that Amazon developed to power its own product recommendation system” [1].

“Amazon says it releases pilots at Amazon Studios periodically for customers to watch and review. Their feedback is taken into account when executives decide which pilots will become a full series. One product of that system is the comedy series *Transparent*, based on a Los Angeles family whose patriarch is transgender. Its debut in 2014 coincided with greater social awareness about transgender issues and was rewarded the following year with the Golden Globe for best TV series, musical or comedy” [8].

Another big media company who uses recommendation engines big time is Netflix. “No one understands the idea of content discovery better than Netflix, because the on-demand streaming video is probably the world’s biggest market for digital consumption of content. Netflix has worked hard to ensure its recommendation algorithms can highlight as much of its content library as possible. In December 2015, Netflix revamped the technology behind its content recommendation engine, deciding to do away with region based preferences in light of their ongoing global expansion” [1].

“Netflix, which distributes shows such as *House of Cards* and *Orange Is the New Black*, pioneered the use of mathematical equations to promote titles that a subscriber might enjoy. That is based on variables such as previously downloaded content, the subscriber’s location and the show’s broader popularity” [8].

“A typical Netflix user may lose interest unless something interesting is found within 60 seconds, two employees of the Los Gatos, California-based company wrote in a paper published in a scholarly journal last year. Netflix’s system for coming up with personalized viewing recommendations helps save more than 1 billion dollar a year by reducing the number of subscription cancellations” [8].

4 CONCLUSION

The rapid growth of big data has given media and entertainment industry an unique opportunity to utilize resources in order to benefit from big data applications and technologies. However, there are still some key challenges media companies are facing such as how to quickly adapt to the big data era, how to deal with and analyze immense amount of data pouring in every minute and how to make cost-effective products and consumer experiences. Examples of current effective big data applications and technologies such as *Match Insights* from SAP and personalized recommendation engines from Amazon and Netflix are provided. In summary, big data applications and technologies are crucial in the success of media and entertainment companies.

ACKNOWLEDGMENTS

The author would like to thank Dr. Gregor von Laszewski for his support and suggestions to write this paper.

REFERENCES

- [1] Shabana Arora. 2016. Recommendation Engines: How Amazon and Netflix Are Winning the Personalization Battle. Web Page. (June 2016). <https://www.martechadvisor.com/articles/customer-experience/recommendation-engines-how-amazon-and-netflix-are-winning-the-personalization-battle/> HID: 233, Accessed: 2017-10-06.
- [2] Lauren Browning. 2015. We sent men to the moon in 1969 on a tiny fraction of the data that's in the average laptop. Web Page. (June 2015). <http://www.businessinsider.com/mind-blowing-growth-and-power-of-big-data-2015-6> HID: 233, Accessed: 2017-10-07.
- [3] Ashok Karania. 2014. How Big Data Is Changing The Entertainment Industry! Web Page. (July 2014). <https://www.linkedin.com/pulse/20140730194648-8949539-how-big-data-is-changing-the-entertainment-industry> HID: 233, Accessed: 2017-10-03.
- [4] Helen Lippell. 2016. *Big Data in the Media and Entertainment Sectors* (1 ed.). Springer International Publishing, Gewerbestrasse 11 CH-6330 Cham (ZG) Switzerland, Chapter 14, 245–259. https://doi.org/10.1007/978-3-319-21569-3_14 HID: 233, Accessed: 2017-10-03.
- [5] Ritesh Mehta. 2017. Big Data in the Field of Entertainment. Web Page. (Aug. 2017). <https://insidebigdata.com/2017/08/20/big-data-field-entertainment/> HID: 233, Accessed: 2017-10-03.
- [6] Tawny Schlieski and Brian David Johnson. 2012. Entertainment in the Age of Big Data. *Proc. IEEE* 100, Special Centennial Issue (May 2012), 1404–1408. <https://doi.org/10.1109/JPROC.2012.2189918> HID: 233, Accessed: 2017-09-20.
- [7] Richard L. Villars, Carl W. Olofson, and Matthew Eastwood. 2011. Big data: What it is and why you should care. *White Paper, IDC* 14 (June 2011). www.tracemyflows.com/uploads/big_data/idc_amd_big_data.whitepaper.pdf HID: 233, Accessed: 2017-09-20.
- [8] Angus Whitley. 2016. How Entertainment Companies Use Big Data. Web Page. (July 2016). <https://www.comstocksmag.com/bloomberg/how-entertainment-companies-use-big-data> HID: 233, Accessed: 2017-10-03.
- [9] Chunjie Zhang, Wenqian Shang, Weiguo Lin, Yongan Li, and Rui Tan. 2017. Opportunities and challenges of TV media in the big data era. In *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*. IEEE, Wuhan, China, 551–553. <https://doi.org/10.1109/ICIS.2017.7960053> HID: 233, Accessed: 2017-09-20.

5 BIBTEX ISSUES