Big Data Analytics in e-Commerce: Understanding Personalisaion

Sohini Chatterjee

Research Department, Shardul Amarchand Mangaldas & Co., New Delhi, India

Abstract: E-commerce has always been guided by the tenet 'know your customer'. The application of big data analytics in e-commerce has enabled the personalisation of shopping to each consumer. Big data sourced from transaction or business activity, click stream, video and voice is processed to micro-segment consumers and derive granular insights. This paper is an exploration into the manner in which such personalisation is done through consumer profiling, predictive analytics, targeted advertisements, improving user experience and price personalisation. Further, it examines accompanying challenges arising from both within and outside the organisation, like organisational culture, derivation of business intelligence, data privacy and security, and data monopolies. The study indicates that while there is considerable cause for optimism about the future of the big data and e-commerce marriage, it is also useful to be cognisant of the limitations of big data and concerns around data privacy and security.

1. Introduction

The e-commerce boom is evident. Global e-commerce sales grew by 13% in 2017 and hit $29 trillion [1]. Some overall drivers of growth are internet penetration, cheap smartphones and mobile data, social media shopping and m-commerce, innovations in supply chain management, enhancements in logistics infrastructure and storage, new digital payment platforms, and big data.

The marriage of e-commerce and big data has forged a virtuous cycle. On one hand, big data has been partly responsible for the prosperity of e-commerce. On the other hand, its prosperity has meant that the newest applications of big data analytics ('BDA') are swiftly implemented in e-commerce. BDA has several uses in e-commerce, such as personalisation, logistics, customer service, and fraud management [2]. This study is limited to studying personalisation, or how the shopping experience is tailored to each individual consumer [3].

The paper aims to study the manner in which big data has helped e-commerce companies to derive business intelligence and personalise their offerings. It is based on the belief that personalisation will be key to the future of e-commerce, with a noticeable move towards increasing personalisation and even hyper-personalisation. Part 2 will set out the significance of knowing the customer in e-commerce. Part 3 will explain big data and BDA. Part 4 will discuss how the marriage of big data and e-commerce leads to personalisation by consumer profiling, predictive analytics, targeted advertisements, better user experience and price personalisation. Part 5 will analyse the challenges of such a marriage, namely organisational culture, derivation of business intelligence, data privacy and security and data monopolies. Part 6 will offer concluding remarks.

2. Know your customer

Retail and e-commerce entities are consumer facing by nature. The nature of their business motivates them to understand their customer and cater in a manner that beats competition. Akin to any customer-centric operation, the quality of service offered by an e-commerce company is hinged on the robustness of its analysis of customer needs and the kind of user experience it delivers. In other words, e-commerce is guided by the cardinal tenet 'know your customer'.

Personalisation in e-commerce is not conceptually novel. However, two developments deserve our attention. First, while personalisation may have been a trend that provides a competitive advantage, it will soon become a standard for every e-commerce company. Failure to personalise will then mean a lack of commercial viability and inability to retain customers. Second, personalisation as a marketing strategy is becoming more refined, with technology imbued modes of assessing needs, tastes and preferences. The e-commerce sector is steadily inching towards hyper-personalisation.

However, the additional difficulty in e-commerce is that the customer is not physically present. The seller cannot simply ask the buyer what she needs. Unlike purchases in brick and mortar stores, online purchases are mediated by technology occurring through a phone, tablet, laptop or vending machine [3].

In light of the above, big data and AI/ML has emerged as an attractive solution. The application of big data and AI/ML have helped e-commerce companies to know their customer and customise offerings. Studies have shown that customers prefer to shop with the same retailer via various channels, and that it is cheaper to retain existing customers than attract new ones [4]. Personalisation is estimated to boost sales by at least 10% and provide five to eight times the return on investment on marketing expenditures [5].

Consider the following illustrative examples:

•Sephora has consistently topped the charts for retail personalisation, due to its focus on the customer experience [6]. The beauty brand boasts of a smooth integration of its online and physical store experiences, a loyalty program that rewards customers with gifts, and a mobile app powered by augmented reality and facial recognition that allows customers to virtually try on make-up.

•Walmart has adopted personalised marketing on its website and mobile app, increasing e-commerce sales by 40% [7]. It introduced a regional trending feature where shoppers were notified of products trending in their own geographical location. The personalised shopping recommendations drew from Walmart's online data as well as transactional data in the physical stores.

•The Indian e-commerce major Flipkart has vouched for "just-in-time personalisation", where every user gets a personalised list of search results [8]. Thus, an identical search query would throw up varying results to different individuals based on their personality and shopping intent at that point in time. Flipkart owned fashion retail unit Myntra has also been working on adding an augmented reality feature to its mobile app. It will use the customer's mobile phone camera to help the them select outfits. It is believed that this will enhance customer engagement.

3. What is the big deal about big data?

Despite its popularity, the term 'big data' does not have a standard definition. Big data predates the Internet and has also been known by other names like business intelligence [9]. It flows from the idea that an organisation should retain data that is a by-product of its operations. Such data may be harnessed to yield insights for improved decision making. As Lin has noted, "the big data hype is simply a rebranding of what many organisations have been doing for a long time" [9]. The evolution has been in the volume of data, kind of data, and analytical techniques applied. Ergo, big data may be viewed as mature business intelligence.

The idea is simple - we can discover more from a large set of information than we can from smaller ones. On the whole, big data is described in terms of the '3 Vs'. A popular definition by Gartner is as follows: "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" [10]. It refers to a vast dataset characterised by the volume of data, variety of types and formats (structured or unstructured), and the velocity of creation [11]. In other words, big data stems from the increasing scope of technology to collect and process fresh predictive information.

Nonetheless, the term big data may be slightly inaccurate [12]. Big data is characterized by much more than its sheer size, and includes other features like variety, velocity and veracity [12]. It is also generally agreed upon that the nature of big data makes it difficult to analyse via traditional modes of data analysis [13].

Big data can be structured or unstructured in nature. The content of structured data centres on demographic information like name, age, gender, residential address, date of birth and preferences [4]. On the other hand, unstructured data includes information from clicks, likes, tweets, videos and voices [4].

However, as will be discussed later in the paper, drawing business intelligence from big data may be an uphill task. AI - and its facilitative tool ML - help in this exercise. The integration of these three concepts can be clubbed together as BDA. While big data refers to technological prowess in collecting and storing vast amounts of data, BDA is the ability to "analyse, understand and gain from the full value of data" [14]. Thus, the absence of BDA would be a dent on big data.

BDA can bring benefits to commerce, society at large and individuals as citizens and consumers [13]. It can enable faster decision making based on data that was previously inaccessible or unusable. In the commercial context, BDA can yield novel business intelligence when used independently or in combination with existing enterprise data. Therefore, it has found use in various sectors like health, transport and e-commerce. As more organisations appreciate the rewards of big data, it has become a priority in the corporate agenda across the world.

4. Big data analytics in e-commerce: the personalisation of online shopping

It has been argued that the buzz around BDA has been generated primarily by web and e-commerce communities [15]. This is not surprising. E-commerce companies understood early on that there is more than one way to listen to the voice of the market. The explosion of the internet and social media thus prompted the collection of large amounts of less structured data that served as a treasure trove of customer behaviour and opinion [15].

E-commerce companies have been one of the fastest groups to adopt BDA, due to their inherent need to stay ahead of competition [4]. Moreover, such competition is based on their ability to understand data. Big data is sourced from transaction or business activity data, click stream data, video data and voice data. Transaction or business activity data is obtained from retail transactions, product consumption and customer grievances gathered in the regular course of business [16]. Click stream data is collected from clicks on social media and online advertisements. While video data is sourced from video in retail settings, voice data is sourced from phone calls, call centers and customer service recordings [16].

Due to the desirability of personalisation as a commercial goal, e-commerce companies attempt to tailor the shopping experience to each customer. As observed above, merely possessing big data may be of limited use. In order to derive business intelligence, e-commerce companies must adopt BDA. The ways in which BDA aids e-commerce companies to personalise the customer experience have been discussed below. All these factors operate in tandem and there may be a degree of overlap.

4.1 Creating rich profiles of consumers

The application of BDA enables e-commerce companies to paint rich and comprehensive profiles of consumers. This is done by accessing both structured and unstructured big data about hundreds of millions of users. Customers provide their personal data while signing up for the website or purchasing an item, such as name, residential address, work address, phone number, age, e-mail address, bank account details etc. This would feed into structured big data.

Moreover, customers may link their e-commerce accounts to their social media profiles. This gives the e-commerce website access to another reserve of the consumer's unstructured data in the form of likes, tweets, videos and voice. Social media profiles are beneficial in evaluating behaviour and sentiment. Data from various channels are seamlessly integrated to produce profiles rich with

demographic and behavioural information. Such integration also unlocks cross-promotional opportunities like discounts on eligible credit cards.

The consumer is tracked on the e-commerce website, related channels and linked social media accounts. The user's product viewing history, purchase history, spending patterns, browsing patterns, clicks etc. give an accurate picture about her shopping preferences. It is common for e-commerce websites to use a combination of session cookies and persistent cookies to identify and track customers. For example, if a user has searched for "puppy costumes" in the past, cookies enable the website to greet the user with more suggestions on puppy costumes.

Thus, an e-commerce company would conservatively know the following information about a consumer: how many times the consumer has visited the website before, which website the consumer visited immediately prior to this, which other websites the consumer visited and for how long, what items the consumer has purchased, what items the consumer has *almost* purchased, the physical location of the consumer, and the computer or browser being used [3]. This may enable companies to know more about the consumers than they know about themselves [17].

4.2 Predictive analytics

Predictive analytics enhance business intelligence and are used by e-commerce companies to predict what a customer is likely to buy next. It is based on an evaluation of the highest price that a customer would be willing to pay for a product, and targeting recommendations accordingly. Big data has enabled the datafication of previously unquantifiable facets like sentiment, intention, friendship and liking [18]. Such datafication is a solution to the need to remain eternally vigilant in pricing. Predictive analytics helps e-commerce companies to exploit price trends and high traffic flows like festivals, by accounting for historical data on products, customer reactions to past pricing, and competitor pricing. Predictive analytics also helps firms in preparing their revenue budgets.

For example, Macy's.com invested in a predictive analytics solution to better understand customers and target their email and web campaigns [19]. It yielded quick dividends in the form of an 8-12% increase in online sales by sending out more targeted emails to registered users and website visitors [19]. Big data has made predictive analytics solutions cheaper and simpler. Predictive analytics also bring additional benefits like fraud minimisation and overall supply chain management.

4.3 Targeted advertisements

Advertisements are increasingly being targeted to customers based on available demographic, geographic and psychographic data. The profiling of consumers enables sellers to categorise consumers into micro-segments with some degree of precision. Advertisers argue that targeting of advertisements is an efficient practice, as consumers are only sent advertisements likely to interest them. For example, ads for sugar free snacks may be targeted to people trying to lose weight.

Targeting is done in various ways. Sellers may adopt algorithms to efficiently micro-target "idiosyncratic consumer preferences with personalised offers" [17]. This may even enable the seller to exploit cognitive limitations (for example, limited will power) of individuals. Connections between a consumer's online and offline behaviour may also be made. Persistent cookies permit sellers to retarget customers who have added items to their cart and not completed a purchase. Demonstrably loyal customers may be shown more advertisements than others. Consumers may also be shown product suggestions, like shoes to go with a pair of jeans. Announcing a marriage as a 'life event' on social media may enable sellers to show targeted advertisements for jewellery in time for a one-year anniversary celebration.

An example of astute targeting is U.S. retailer Target's program to predict early stage pregnancy by tracking slight changes in purchases (like buying nutritional supplements and unscented lotion) [20]. This allowed the store to display targeted ads and attract expecting mothers [20].

Moreover, sellers no longer need to wait for the consumer to visit. Consumers may be targeted with advertisements at any point of time in the day via multiple feeds [3]. This trend will be amplified with the pervasion of Internet of Things connecting daily objects and making them ripe for advertising.

4.4 Better user experience

BDA, predictive analytics and behavioural analysis help e-commerce companies to use real time analytics and design personalised offers, recommendations and promotions for customers. For example, data on the uptake of a campaign with various audience segments may be useful in identifying loyal customers [16]. The seller can then construct a more interactive and engaged customer experience, which could translate into increased sales. A visible trend may be observed in the creation of subscription-based loyalty programs that provide perks to members. For example, BDA can help an e-commerce company in identifying a particular customer's liking for high end mobile phones and target a subscription-based membership package to her.

4.5 Price personalisation

BDA allows e-commerce companies to deliver personalised prices to customers. In other words, it can engage in price discrimination based on an assessment of the customer's preference and tolerance. Sellers can increase revenues by quoting a higher price to customers willing and able to pay more. One study found that several retailers and travel websites mark personalised prices that can differ by as much as hundreds of dollars from one consumer to the next [21]. This is often done through algorithmic price optimisation. Such price discrimination may consider information like the consumer's location, the time of day, the characteristics of their computer and their spending history and patterns [21].

5. Challenges to the use of big data in e-commerce

The discussion above has revealed that big data permits e-commerce companies to personalise in myriad ways. However, despite its usefulness, there may be significant challenges to exploiting it fully. Four such challenges are discussed below:

5.1 Organisational culture

Cultural transformations do not happen overnight. Established organisations that have heavily invested in big data have found it difficult to shift to a "data-driven culture" within the organisation [22]. Key organisational challenges have been organisational alignment, resistance, and change management. One survey has revealed that 99% respondents said that while their firms are trying to become data-driven, only a third have succeeded [22]. To make good of big data investments, companies may need to make big data understandable to employees at the fore. Employees may not trust big data or BDA insights, and managers would have to present it in a comprehensible form, say through visualisation [4]. In this regard, start-ups and small and medium enterprises (SMEs) may enjoy an advantage as they enjoy the flexibility to create a data-driven culture at the outset.

5.2 Derivation of business intelligence

The big data abundance has still not answered how to derive meaningful business intelligence. An average company uses less than 10% of data they collect and store [23]. This makes them data rich and insight poor, leading to a loss of revenue from the costs of collecting and storing data. BDA and data mining are required to extract nuggets of business intelligence from the noise. Such nuggets would contain highly actionable information like how to increase customer engagement or customer loyalty [23].

Big data generally succeeds in capturing what is easy to capture [12]. It is not meant to deliver deep insight on anything in particular. Big data assists in answering 'what' and not 'why'. It would be glib to suppose that big data provides an omniscient and unerring picture of the world. To use an apt analogy, big data is akin to extracting nuggets through open pit mining [12].

Moreover, big data is limited by sampling frame, the technology deployed, the context of data generation, and how the data has been classified [12]. Insights from big data will only be as good as the data it is originally fed. Big datasets containing biased and flawed data will produce inaccurate or invalid insights [12]. For example, BDA insights generated from tweets by fake accounts, hacked accounts and bots may yield invalid insights [12]. Doubts have also been raised on the reliance on

social media profiles as an indication of what a customer likes. Further - correlation may be mistaken as causation - as big data is merely trying to identify the association between things rather than the deeper reasons underlying it [18].

This messiness appears to be an inherent part of big data [18]. Thus, some scholars have cautioned against "big data hubris" - a tendency to view big data as an ultimate solution [9].

5.3 Data privacy and security

Big data may consist of non-personal data (for example, climate, weather and pollution data) and personal data. Where big data is also personal data, it may clash with privacy and data protection. Big data may potentially worsen the power imbalance between an individual and a corporation, and an individual and the State [18].

Prof. Calo has argued that personalisation takes market manipulation to the next level [3]. Previously unquantifiable aspects of individual behaviour and personal lives are now being datafied [18]. Insights from datafication are being used to personalise e-commerce and boost revenue. Further, big data may be combined with proprietary data and open data, leading to potential identification of individuals. Thus, the expanding frontiers of personalisation may constitute a privacy threat.

Privacy has several dimensions, including the right to be left alone, control over one's personal information, dignity, autonomy and freedom. In the domain of informational privacy, it is accepted that consent should be free, informed, specific and clear. However, dense privacy policies, built in dark patterns, one-sided cookie consent notices and consent fatigue render consent meaningless. Therefore, how much control users exert over the personal data they share with sellers is dubitable. Further, micro-segmenting of consumers also raises group privacy concerns. While practices like first degree price discrimination maximise commercial revenues, it has an adverse effect on transparency for consumers.

Personalisation also raises data security concerns about how secure the collected and stored data is. This is especially relevant considering that such data may be shared within the organisation among multiple concerns, and shared outside with third party entities like advertisers and data brokers. Both the privacy and security issues are interlinked, as privacy also demands that personal data is securely stored and transmitted within and outside the organisation [24].

BDA equips firms to exploit irrationalities and vulnerabilities of individuals which may lead to actual and perceived harms [3]. Prof. Calo has noted that big data presents a new and more efficient way to "identify cognitive bias by attempting to pinpoint profitable anomalies" [3]. For instance, an e-commerce company may target an obese person with donut ads, knowing enough about her weaknesses.

Additionally, e-commerce concerns may put their big data to research and development uses. Ethical issues with respect to privacy and security may go unaddressed as corporate big data practices are not subject to ethical oversight unlike a university conducting research on similar lines [25]. Corporate big data is inaccessible and thus the likelihood of an unbiased privacy impact assessment is doubtful.

With the development of data protection laws across the world, the coming into effect of the EU GDPR, and the framing of data subject rights like the right to be forgotten, e-commerce companies should take data privacy and security seriously. In this regard, some scholars have argued for differential privacy protection or privacy customisation, where a user with a lower demand for privacy protection is permitted to waive her privacy in return for the right products [26]. It has also been suggested that privacy disclosures should be personalised as different groups may view privacy risks differently [27].

Such proposals are worth considering in the context of e-commerce, as long as it is underpinned by informed consent. A failure to address privacy and security concerns may negatively affect consumer perceptions towards e-commerce and distress the sector.

5.4 Data monopolies

Some companies are privileged actors in their access to big data by virtue of their large user base. E-commerce behemoths have gigantic reserves of big data to spur innovation [28]. These companies have been collecting and analysing data for years and enjoy a head start over newer players. Further, with the trend to become one-stop-shops, competition concerns arise regarding abuse of dominance. Thus, there is a divide between data rich and data poor players in the market.

Such dominance may distinctly disadvantage start-ups and SMEs. While it is also possible that investment of big companies may unlock new opportunities for start-ups (like Uber's strategy of re-harvesting Google Maps data), it cannot be denied that newer organisations are likely to face a steep learning and adoption curve. Moreover, they would only have access to big data in the public domain. On the other hand, established e-commerce players will have more resources to design large data sets and the requisite infrastructure to fully exploit that data [28].

6. Conclusion

Information and communication technologies have transformed the capacity of e-commerce companies to offer personalised services. BDA has proved to be immensely valuable in maturing business intelligence and telling entities how to make better business decisions.

This paper has explored the ways in which BDA has enabled the personalisation of online shopping. The construction of rich consumer profiles replete with demographic, geographic and psychographic information has made it possible to improve the user experience and engage in predictive analytics, targeted advertisements, and price personalisation. E-commerce companies are in a firm position to know the customer and give her what she wants.

The paper has also studied some attendant challenges to the deployment of BDA arising from both inside and outside the organisation. A review of internal challenges indicate that it is wise to avoid big data hubris, as big data is suited more for information rather than explanation [18]. Big data suffers from inherent limitations that may come in the way of deriving meaningful business intelligence. It is futile for an organisation to heavily invest in data analytics and end up being data rich and insight poor.

In external challenges, data privacy and security, and the danger of data dominance are noteworthy. With more facets of individual behaviour being datafied, big data has the potential to tip the scales of power against the individual. Further, access to big data may create a divide between data rich and data poor players in online commerce. Effectively tackling these challenges will shape the future of e-commerce and the gains it can bring to consumers.

Acknowledgement

References

[1] United Nations Conference on Trade and Development, Press Release: Global e-Commerce sales surged to $29 trillion, 2019.

[2] Balaraman, P., Chandrasekar, S. E-commerce trends and future analytics tools. Indian Journal of Science and Technology, 2016, 9 (32).

[3] Calo, R. (2014) Digital market manipulation. The George Washington Law Review, 82(4), 995-1051.

[4] Akter, S., Wamba, S.F. Big data analytics in e-commerce: a systematic review and agenda for future research. Electronic Markets, 2016: 173-194.

[5] Liebowitz, J. Big data and business analytics. CRC Press, 2013.

[6] Sailthru. Personalisation Index Retail, 2018.

[7] Sailthru. Deep Dive into the Retail Personalisation Index Top 10: Walmart.

[8] Flipkart Tech Blog. Just In Time Personalisation, 2018.

[9] Lin, J. On building better mousetraps and understanding the human condition: reflections on big data in the social sciences. The Annals of the American Academy of Political and Social Science, 2015, 659: 33-47.

[10] Gartner, IT Glossary: Big Data.

[11] Porche, I.R., Wilson, B., Johnson, E., Tierney, S., Saltzman, E. Big data: challenges and opportunities, data flood. RAND Corporation, 2014.

[12] Kitchin, R., Lauriault, T.P. Small data in the era of big data. GeoJournal, 80(4): 463-475, 2015.

[13] Information Commissioner's Office, UK, Big Data, Artificial Intelligence, Machine Learning and Data Protection.

[14] European Data Protection Supervisor. Opinion 7/2015: Meeting the Challenges of Big Data- A Call for Transparency, User Control, Data Protection by Design and Accountability, 2015.

[15] Chen, H., Chiang, R.H.L., Storey, V.C. Business intelligence and analytics: from big data to big impact. MIS Quarterly, 2012, 36(4), 1165-1188.

[16] Stock, D. Driving e-commerce growth using big data analytics. NOVA Marketing Insights, 2018.

[17] Wagner, G., Eidenmüller, H. Down by Algorithms? siphoning rents, exploiting biases, and shaping preferences. The University of Chicago Law Review, 2019, 86(2), 581-609.

[18] Cukier, K., Mayer-Schoenberger, V. The rise of big data: how it's changing the way we think about the world. Foreign Affairs,2013, 92(3), 28-40.

[19] SAP. Macy's.com Gains Better Insight into Customers with Predictive Analytics from SAP, 2014.

[20] McGregor, V.K., Calderón, S.H., Tonelli, R.D., Big data and consumer financial information. Business Law Today, 2013.

[21] Hannak, A. et al. Measuring price discrimination and steering on e-commerce web sites. Proceedings of the 2014 Conference on Internet Measurement Conference, 2014: 305-316.

[22] Bean, R. How companies say they're using big data. 2017.

[23] Bisack III, J.H. Business intelligence: finding nuggets in the noise. The Journal of Private Equity, 20(3), 9-11, 2017.

[24] Miyazaki A.D., Fernandez Ana. Internet Privacy and security: an examination of online retailer disclosures. Journal of Public Policy & Marketing, 2000, 19(1), 54-61.

[25] Richterich A. Big data: ethical debates. The Big Data Agenda, University of Westminster Press, 2018.

[26] Zhou, W., Piramuthu S. Information Relevance Model of Customised Privacy for IoT, Journal of Business Ethics, 2015, 131(1): 19-30.

[27] Busch, C. (2019). Implementing personalised law. The University of Chicago Law Review, Symposium: Personalised Law, 86(2): 309-332.

[28] Wessels B. et al. Big data, open data and the commercial sector. Open Data and the Knowledge Society, Amsterdam University Press, 2017.