Economics of big data: review of best papers for January 2018

Verstappen, Ksenia

Research Consulting and Development

February 2018

Economics of big data: review of best papers for January 2018

Ksenia Verstappen*

RCD, Kyiv, Ukraine

ABSTRACT

Hundreds of new papers on big data are released every month and at times it is difficult to distinguish between them in terms of quality and practical use. The purpose of this monthly review is to highlight the findings in the most relevant papers in Economics of big data to help readers identify the most important new developments in the field. The review for January 2018 includes a study of social networks in truancy, a paper on consumer privacy and data collection and three NBER papers on applications of Artificial Intelligence in Economics.

INTRODUCTION

The review of best new papers in Economics of Big Data for January 2018 includes a study of social networks in truancy, a paper on consumer privacy and data collection and three NBER papers on applications of Artificial Intelligence in Economics. These best papers were chosen based on a two-stage selection process: a short list of the papers was identified by the editor while the author further refined the list to include only 5 best papers.

SOCIAL NETWORKS IN TRUANCY AND THE TARGETING OF TREATMENT

Researching the links between social networks and truancy as one of types of risky behaviors of the young people, Bennett and Bergman (2018) demonstrate how the incentives to school absenteeism may be attributed to the social networks connections between students and offer a possible intervention to address truancy that involves the development and implementation of an optimal-targeting algorithm. Basing on the fact that in most cases, the students tend to miss separate classes rather than whole days and hypothesising that they often do it with some of their friends or acquaintances, the scholars offer the solution that creates a spill-over effect across the students' network. They demonstrate how the school's administrative data on truancy allows to reconstruct the absentees' social networks (in other words, what students miss classes with who), and how text alerts to the parents that are sent each time their child misses a class may create the spill-over effect to other students within this social network. The study demonstrates that although it can never be stated with full certainty that the students miss classes together, the number of joint cases of truancy is much more frequent than it could be expected by chance. The method of sending automated texts to the parents of the students who miss classes demonstrates significant spill-over effect, giving incentives to other students within the same social network to

attend the school in a due manner. The results of the study also suggest the importance of leveraging the students' social network: e.g., in order to make the intervention more effective, the best way is to apply the treatment (a text message sent to the parents) to those students who do not demonstrate truancy on a regular basis, and yet whose strongest connection in a social network is someone who is chronically absent. Overall, the article identifies the social network effects and demonstrates how the offered intervention may benefit the families, schools, and society at large, serving as a cost-effective method of preventing truancy.

PRIVACY AND PERSONAL DATA COLLECTION WITH INFORMATION EXTERNALITIES

Privacy concerns being one of the most topical issues for the public when it comes to collecting, storing and monetizing personal data of the Internet users, there is a need to develop effective policies that would allow to bridge the gap between protecting individual users' privacy and the growing activity of data brokers. Choiy et al. (2018) demonstrate how the model of privacy functions within the market equilibrium on the example of a sample monopoly website selling content and collecting the data about the clients only upon their consent, the consequences of which should be fully known to the users. This model demonstrates that the loss of consumer privacy under such circumstances is much higher compared to the social optimum, even in case the users are informed about the consequences of agreeing to provide their personal data.

One of the worrying consequences of users' careless behaviour concerning their own personal data is information externalities. The users who did not agree to share their data may lose their privacy due to the improving ability of big data analytics to make predictions and draw conclusions about such users basing on the personal data provided by other people. This may have negative social influences, such as deterring the creation of greater data networks. The authors also research the implications of the market structure where various small websites may not

intend to get access to users' personal data, however, the emerging data brokerage firms that are willing to purchase this data with the goal to aggregate it and use may serve as an incentive for the small websites to start personal data collection.

The authors show how this model may also facilitate the emergence of negative information externalities which, in their turn, can result in the increased number of websites entering the market to collect and sell personal data to the brokers, and eventually the consumers will find themselves in less favourable circumstances in the market equilibrium. The researchers suggest that the improved privacy regulations are needed in order to address the problem of growing data broker industry for the sake of public good.

AI AND INTERNATIONAL TRADE

Tackling the problem of the relationship between Artificial Intelligence and the functioning of markets, Goldfarb and Trefler (2018) state that there is still little awareness when it comes to the impact of AI on the international trade. With the current situation of uncertainty when international agreements are being negotiated which would put constrains upon government regulations of AI, in-depth research is required in order to understand what policies and regulations need to be developed.

The scholars research how Artificial Intelligence interacts with several major dimensions of the international trade, including competition, knowledge creation/diffusion, and scale. On the example of China whose current economic success is largely attributed to the fact that its protection laws do not allow international companies to penetrate its market, while the country's own commercial AI companies are flourishing and will most likely soon become globally recognized. This example demonstrates how behind-the-border regulations may eventually shift the balance of global power. The scholars argue that the current U.S. policies that are based on the "zero sum game" cannot be considered the right approach when it comes to further development of the Artificial Intelligence

policies.

While trade economists have developed a number of effective policies allowing to promote an industry, certain features of Artificial Intelligence (particularly knowledge creation/diffusion, scale, and competition) may affect the appropriateness of the existing policies. This creates an incentive to come up with the new policies and regulations. While suggesting that overall, cluster policies are likely to yield satisfactory results when applied to the Artificial Intelligence market, there are several difficulties arising from the nature of this specific market (international knowledge diffusion and data scale). There are also certain regulatory problems that may affect international trade policies, one of them being access to data. The authors identify five of such policies (rules of data localization, domestic privacy policy, access to the data of governments, protection of source code, and industrial regulation in the Artificial Intelligence application spheres), suggesting their implications for international trade.

To summarize, the authors emphasize on the implications of the Artificial Intelligence for international macroeconomics and suggest that further research into the nature of the technology (and particularly, knowledge creation/diffusion, scale, and competition) is required in order to develop the trade models that will be useful and effective in the long run and on the global scale.

ARTIFICIAL INTELLIGENCE AND CONSUMER PRIVACY.

Ginger Zhe Jin (2018) tackles the problem of breaching consumer privacy and insufficient data security during the big data era, due to the opportunities that have become available for creating innovations in the sphere of artificial intelligence. The researcher points out the growing information asymmetry when it comes to consumer data collection, processing, and use; particularly when it comes to future data use whose consequences may still be unknown to the consumers. Some of the most evident problems in this area are:

1) The vagueness of the concept of future data use that may evolve with the time to fit the purposes of the seller or data brokers, while remaining out of the consumers' control.

2) Externality of personal data collection that manifests itself in the fact that while most of the benefits from using the data are collected by the sellers, any possible misuse of data in the future by the third parties (such as fraud or identity theft) would mostly affect the consumers.

3) Commitment issue arising from the fact that it is extremely difficult, if not impossible, to trace any damages caused to the consumers to the data collectors or prove the existence of any such link. This may serve as an incentive for the sellers to promise to the clients to adhere to a certain consumer-friendly data collection policy, and renege in the future after the data has been collected.

These three problems are likely to be further aggravated once AI comes into action, encouraging the sellers and third parties use the data intrusively, creating even more risks for the consumers. The scholar questions the existing data protection laws and calls for further research in various areas, from computer science, information science and economics to law, marketing, and statistics, so that to restore the companies' accountability for the risk they may cause to the security of consumer data.

THE TECHNOLOGICAL ELEMENTS OF ARTIFICIAL INTELLIGENCE.

Matt Taddy (2018) offers the framework for regarding the structure of Artificial Intelligence that combines various Machine Learning algorithms. This framework can be effectively used by businesses working in the field of Artificial Intelligence. The scholar singles out the following elements of the Artificial Intelligence as the system able to perceive the human-level knowledge and automate the processes that had been performed solely by humans before:

1) Domain structure (business/economic domain

expertise) that makes it possible to divide complex tasks into smaller ones that can be processed with the help of Machine learning. While the scholar illustrates this element using the example of a computer game, he stresses that the ability to break the complex human problems into the smaller tasks that can be solved using Machine Learning is one of the key success factors for using Artificial Intelligence in business.

2) Bank of data plus the strategy of data generation. The author suggests that "generation" is a better term here than the more common "collection," because when it comes to the systems of Artificial Intelligence, the stream of information needs to be consistent and up-to-date, which calls for an active data generation strategy. This element is extremely important for business, since large volumes of various data are needed for making the Artificial Intelligence work.

3) Machine Learning processes and algorithms that are able to automatically make predictions basing on the complex unstructured data. This part of Artificial Intelligence, Taddy argues, if often confused with the AI itself, which is a common misconception, and he explains and illustrates the differences between the two concepts. Currently, Machine Learning is successfully used by the businesses for gaining important insights from the data using business analytics models.

Having singled out the key components of Artificial Intelligence, the article helps understand how Artificial Intelligence fits into the picture of big data analytics, and how it can be used by businesses.

REFERENCES

Bennett, Magdalena and Bergman, Peter. 2018. "Better Together? Social Networks in Truancy and the Targeting of Treatment." CESifo Working Papers, 6848.

Choiy, Jay Pil, Jeon, Doh-Shin, and Kim, Byung-Cheol. 2018. "Privacy and Personal Data Collection with Information Externalities." Working Papers, 18-887.

Goldfarb, Avi and Trefler, Daniel. 2018. "AI and International Trade." NBER Working Paper Series, 24254.

Jin, Ginger Zhe. 2018. "Artificial Intelligence and Consumer Privacy." NBER Working Paper No. 24253.

Taddy, Matt. 2018. "The Technological Elements of Artificial Intelligence." NBER Working Paper Series, 24301.