From data to context-aware decision making: challenges and opportunities

Kerstin Bach

Department of Computer Science
Norwegian University of Science and Technology, Trondheim, Norway
http://www.idi.ntnu.no

Abstract. This extended abstract presents the research challenges for developing context-aware, explainable artificial intelligence challenges and briefly presents the opportunities ahead when utilizing the vast amount of data generated.

Keywords: Big Data, Context-Awareness, Explainability, Decision Support Systems

1 Introduction

The potential of context-aware applications driven by big data applications providing tailored information when needed is one of the challenges our society that is facing in the near future. As of today almost every activity we take can be monitored, analyzed and processed, but putting these information in a context and providing personalized information, advise or content is still challenging.

Having vast amounts of data available does not necessary mean we can automatically derive a single users preference. Before we sketch out the challenges for personalization in big data applications, we will introduce the most commonly used definition to describe the term big data is the one by Gartner (the 3 Vs)[1] :

> Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.

While high-volume is describing the growth in terms of data points and according to a Gartner study[2], the information volume worldwide is growing at a minimum rate of 59 percent per year. Eventually, this will lead to much more produced data volumes than available storage – but more importantly bottlenecks in data analytics will emerge. The second aspect is the high variety of data types which does not come ready for processing nor integration, because media types, dimensions and representations differ. The probably hardest aspect in handling big data is the velocity of the data stream that sets the pace in

[1] http://www.gartner.com/it-glossary/big-data/
[2] http://www.gartner.com/newsroom/id/1731916

which data has to be analyzed and processed, but challenges applications for synchronizations or coherence in data processing.

The second aspect targets personalization, which is according to Mulvenna et al. [3] defined as follows:

> Personalization is the provision to the individual of tailored products, services, information or information relating to products or service.

Personalization adds up the the 3Vs since it requires to process data and customize it to the individual, who's preferences might be changing. Not only, that the vast amount of data has to be processed, it also has to be fitted in the context and needs of a user. Therefore the combination and fusion of data for personalization requires novel models that adapt and learn over time.

Eventually all aforementioned aspects will feed into decision making. Therefore the decision making process can by carried out (a) by a system, (b) by a human after reviewing analyzed data or (c) by a human based on a recommendation (decision support systems). The topic of decision making and artificial intelligence has been addressed over the years (such as [4]), however the field in which it is applied today is limited.

2 Significant Research Challenges

2.1 Deriving Value from Data

Today we are aware that we produce more data than we are capable of storing and processing, hence we have to select the right data to analyze and the right content to present a user. Moreover, we have to find ways to create useful information rather than interesting information. The key in research should be in creating methodologies on selecting, analyzing, interpreting and guiding the usage of data. For example streaming applications such as Spotify[3] or Netflix[4]are able to answer questions based on their statistics (e.g. Which was the most played song in 2014?) easily, however it gets complicated for an application to

- predict what a user would like to do/buy/etc. next, or
- manipulate the user (to click an ad/buy a product/etc.)

Big data empowers applications to become more personalized and companies to get to know their customers better. However, in order to build and use this knowledge, we have to work on methodologies that include an overall view on how to select and utilize data sources in order to create actionable suggestions. There is a lot of work carried out on handling big data, but when there is a certain problem to solve, it is extremely challenging to find the right data along with methods to solve given problem. Further on, a lot of focus went into analyzing and presenting the past to the user, while we claim that this information should be used to guide to user in the future through predictions.

[3] https://www.spotify.com
[4] https://www.netflix.com

2.2 Data-Driven Decision Making

Since we can collect a vast amount of data at unprecedented speed, we have to investigate how to utilize it for understanding the past, predicting or even changing future outcomes . It is not the velocity that creates the biggest challenge, it is the speed of the feedback loop, taking data from input through to decision, that is important. Creating meaningful, non-trivial recommendations or providing insight on relevant questions goes beyond what personalization allows today. We just started to capture information about a user, but understanding their behavior and and incorporating social signals from all available data sources can revolutionize the way we build applications.

2.3 Data Analysis and Prediction

Applying statistical data analysis on heterogeneous data sources for predicting certain outcomes is challenging and even though technologies are at hand, they might not fit in a big data setting, because the incoming data stream is too volatile or the data quality is not sufficient. Predictions always required certainty, while in real-world applications data sets tend to be incomplete and hence predictions are made under uncertainty.

Challenges arise from transforming existing machine learning algorithms into the aforementioned setting allowing them to perform on a high-scale.

2.4 Explainability

Creating systems that do not only provide a recommendation, but also an explanation to the user is key to increase the trustworthiness in system. This movement is supported by the requirements set by the upcoming General Data Protection Regulation (GDPR) by the European Union as well as by the strategic DARPA program on Explainable AI (XAI)[5]. Especially GDPR has recently triggered an important discussion on the "right to explanation" initated by Goodman and Flaxman [2] and the response by Wachter et al.[5].

3 Opportunities

Big data or better smart, context-aware data is a term that describes the challenges for applications: making use of the data we can obtain in a smart way. It is not necessarily solved by collecting more data solves, unless applications can make a use of it by creating something its user wants or needs. How can applications be smarter then? Using contextual information for personalization is certainly a major step forward, but this has to be combined with feedback from the user and the ability to react on it when necessary.

In the following we will point out possible opportunities for developing context-aware, affective computing that should be targeted in order to develop the field forward.

[5] https://www.darpa.mil/program/explainable-artificial-intelligence

3.1 Value Creation

Providing personalized information to a user means on the one hand understanding a user's needs as well as having necessary sources for creating relevant content. Therefore we have to analyze the past, observe the present and provide interesting content and recommendations for the future. The value within an application highly depends on its context and being able to automatically collect and use contextual information will lead to better applications.

Contextual information can make a service adaptive to a current situation or create content in a way it can be consumed by the user. Also knowing in which setting an application is used should influence the type of presentation, detail and complexity. We certainly have to differentiate whether a user is willing to spend time in an application, because then the information presented can get more complex, stories can be told and knowledge can be transferred. On the other hand, when users are only consuming an information with a short attention span, the message has to be short.

Interesting research questions are arising when creating valuable content:

- Which type of data is required in the background?
- How can neccessary data be identified and/or acquired?
- How to select the right data stream for a service?
- How to measure the quality of the results?

3.2 Real-time decision making

As mentioned before closing the feedback loop between users and recommendations is challenging, since systems have to react on a user's behavior in real-time. Therefore, it is important to have a true assessment of the context in order to describe what is going on right now and enabling systems to provide immediate feedback on decision.

Decision making can be seen in two ways: machine and human centered. For machines it describes the automatic selection of actions, e.g. which content to be provided or which ad to be posted. This also includes follow- ing up the results on the decision and its comparison to set expectations. For humans it means providing suggestions based on data analysis and observing certain behaviors or sensors. Big data applications are able to take more features into considerations as well as they can share experiences captured in data.

3.3 Methodologies

As of today, there are many tools and platforms available that process big data, but there is certainly a lack of how to find the right set-up for addressing a given problem. The big data landscape[6] shows the high variety of software, which makes a selection for a new application challenging. Most technologies

[6] `http://www.bigdatahadoop.info/understanding-big-data-ecosystem/`

listed have their own purpose and an individual application is built out of this toolboxes.

Data is often available, because companies have learned that there is a huge amount of value in their data. However, discovering and materializing the sources is very challenging and hence often described as partially science, partially art. Breaking relevant tasks down by creating guidelines will help less data-driven companies getting grip on their data and leverage innovation. Hence goal should be the following:

– How to can existing data be used?
– What is required for data-driven innovation?

3.4 Rapid prototyping

One aspect of big data applications is the source data, but on the other hand prototypical applications and proof-of-concepts should be developed to undermine hypotheses. Hence guidelines or methodologies have to be showcased in applications, open to others in order to provide insight in big data applications.

Those prototypes should not be products, but they necessarily have to work with real-data to be convincing. Along with the prototypes itself, their dissemination is crucial, so the results of conducted research can be applied.

4 Conclusion

Big Data Analytics describes a value within our today's society and bridging between the findings in research and the insights and needs of an industrial setting is challenging. Hence a collaboration, where both theoretical approaches and prototypical applications can be merged, is fruitful for academic research, but also for industrial collaboration.

Moreover, our recent work on developing decision support systems for patients in primary care (see [1]), public health [7] or fish farming[8]. In the context of those projects, developing context-aware and explainable AI systems is a core focus.

References

1. Bach, K., Szczepanski, T., Aamodt, A., Gundersen, O.E., Mork, P.J.: Case representation and similarity assessment in the selfback decision support system. In: Case-Based Reasoning Research and Development - 24th International Conference, ICCBR 2016, Atlanta, GA, USA, October 31 - November 2, 2016, Proceedings. pp. 32–46 (2016)
2. Goodman, B., Flaxman, S.: European Union regulations on algorithmic decision-making and a "right to explanation" (Jun 2016), arXiv e-prints: 1606.08813

[7] https://www.ntnu.edu/hunt4
[8] http://exposedaquaculture.no/

3. Mulvenna, M.D., Anand, S.S., Büchner, A.G.: Personalization on the net using web mining: Introduction. Commun. ACM 43(8), 122–125 (Aug 2000), http://doi.acm.org/10.1145/345124.345165
4. Pomerol, J.C.: Artificial intelligence and human decision making. European Journal of Operational Research 99(1), 3–25 (1997), https://EconPapers.repec.org/RePEc:eee:ejores:v:99:y:1997:i:1:p:3-25
5. Wachter, S., Mittelstadt, B., Floridi, L.: Why a right to explanation of automated decision-making does not exist in the general data protection regulation. International data privacy law. 7(2), 76–99 (2017)