On Data Driven Organizations and the Necessity of Interpretable Models

Tony Lindgren(✉)

Department of Computer and System Sciences, Stockholm University, Borgarfjordsgatan 12, 164 40 Kista, Sweden
tony@dsv.su.se

Abstract. It this paper we investigate data driven organizations in the context of predictive models, we also reflect on the need for interpretability of the predictive models in such a context. By investigating a specific use-case, the maintenance offer from a heavy truck manufacturer, we explore their current situation trying to identify areas that needs change in order to go from the current situation towards a more data driven and agile maintenance offer. The suggestions for improvements are captured in a proposed data driven framework for this type of business. The aim of the paper is that the suggested framework can inspire and start further discussions and investigations into the best practices for creating a data driven organization, in businesses facing similar challenges as in the presented use-case.

Keywords: Data driven framework · Interpretability · Organization

1 Introduction

As organizations such as governments, corporations, educational institutions (from here on referred to as organizations), start to re-organize their work with the intent to harness and utilize the data that they are (or could be) in possession of, they are confronted with different challenges. One can partition the challenges into two major themes, technical and organizational. The technical challenges can include how to efficiently collect data, how to store it, how to query and analyze it etc. On the organizational level decisions on which department will do what and their responsibiltes must be made. Typically, the end goal for an organization that have initiated change in the above mentioned direction is to "become data driven". The silent assumption is that a data driven organization is the pinnacle of efficiency, with very low waste, be it in time, effort or any other scares resource [10].

One organizational model that subscribes to this data driven organizational idea is the New Public Management (NPM) [2], which strives to generate feedback loops which can be used to guide the organization. These feedback loops have become more common everywhere in society. It can materialize itself in

A.-S. K. Pathan et al. (Eds.): SGIoT 2018, LNICST 256, pp. 121–130, 2019.
https://doi.org/10.1007/978-3-030-05928-6_12

the form of survey machines at the supermarket or after you rented a car you will receive an e-mail or text-message from a survey firm with questions. Some organizations have this feedback loop tightly coupled to their business model, for example Airbnb and Uber, while others like Facebook and Google tries to get feedback from your GPS data to be able to make better recommendations in the future. In Fig. 1 two examples of creating feedback is shown. There have been quite a few attempts to formalize the activities and the different steps needed to successfully create data driven organizations, these descriptions come at different abstraction levels, from practical and concrete descriptions of the process steps needed to facilitate analysis of data [9] to suggestions on how to organize work [10]. One aspect of becoming a data driven organization is the acceptance and utilization of abstractions and inferences made from the data provided both by the explicit feedback-loop and other data relevant to the service provided by the organization. Data science and data mining is the study of how to, from data, create such abstractions which typically involve the process of searching for patterns or regularities in data and then describing these regularities through a predictive model. Predictive models have been put into use by organizations more frequently in recent years, and is one of the reasons for the increased interest in the area of model interpretability. For a good overview of this area the reader is encouraged to read [7]. In this paper we are interested in the interpretability of inductive models and how this feature fit into the concept of data driven organizations.

(a) (b)

Fig. 1. A few examples for creating a feedback loop

The rest of the paper is structured as follows: first we will look into the lifecycles of inductive models in data driven organizations, then we will look into the role of interpretability for inductive models. We will then present our use-case and its problems then present a data driven framework which addresses these problems. Finally, we finish off the paper with a discussion about the proposed framework.

2 Life Cycle of Inductive Models in Data Driven Organizations

One of the most commonly referenced process for doing data mining is the *CRoss-Industry Standard Process for Data Mining* (CRISP-DM) [9]. CRISP-DM divides the data mining process into six different phases. The phases are: business understanding, data understanding, data preparation, modelling, (model) evaluation, and (model) deployment. Business understanding include the major steps of understanding the business objectives for initializing the data mining effort and converting this knowledge into a data mining problem and designing a plan to reach the business objectives. Data understanding, data preparation and modelling are all phases which are instrumental to the specific data mining task. Model evaluation are important as this step checks whether or not the business objectives are fulfilled by the inductive model. The next step is to proceed into the deployment phase, if the objectives are fulfilled. CRISP-DM does not touch upon how work should or could be organized to facilitate the suggested process phases. One can also argue that CRISP-DM put very limited emphasis on the steps after deployment.

Traditionally organizations have been organized in a hierarchical fashion, where different parallel organizations have had different functions, be it research and development, production, aftermarket, customer relations etc. Still this way of organizing work is common. Difficulties which have been identified with such a structure is that such organizations can be slow to react upon changes in its environment, and that cross functional work can be hindered [1]. This has led to the advent of organizations which focus on the process or in values streams to make them explicit and assign the responsibility of certain parts of the process to certain parts in the functional organization [3].

In their article [6], which focus on humans working in data driven organizations, and the impact this has on their working environment. The article investigates two ridesharing organizations, Uber and Lyft, in which the human jobs are assigned and evaluated by algorithms in a data driven framework. Both Lyft and Uber allow for few managers to, via algorithms manage thousands of drivers worldwide. In both companies the drivers log in to an application, typically on a mobile device, to signal that they are ready to receive jobs. Exactly how the algorithm matches vehicles and customers is secret in both cases but proximity is one of the factors.

Active drivers have a limited time (15 s in case of Uber) to accept a fare or not. Areas with many customers and few vehicles are signaled in the app, to encouraged drivers to go to this areas, as incentive drivers will receive a higher compensation (pay) for these fares. The drivers can be judged by their customers and the driver can also rate the passengers. The main feedback loop into the data driven system is the ratings and for drivers the percentage of accepted drives w.r.t. offered fares. The drivers also get payment promotions (each hour) if they are active for longer periods. All these variables can cause drivers, who have a good understanding of how the algorithm works to, for example, park between two other uber-cars while having lunch to make sure that they are still online

but have a small risk of actually receiving a fare. Yet another consequence is that some drivers who don't want to take customers from a certain area turn off their application while passing through that area, to avoid having to pass on a fare.

The main findings in the study indicate that drivers were to some extent frustrated by the lack of transparency of why they were offered some fares even though they were not the closest, or being judged when they passed up fare even though they had a good reason for it. The authors of the study suggest that improving the transparency of the algorithm most probably would increase the acceptance and trust of the fare assignments, but also might come at the cost that more drivers will "know" how to exploit the algorithm.

Another type of data driven management is NPM, this has received criticism for just creating more bureaucracy and not have the good effects on steering the organization which was the original idea. Hoggett [4] provide one description of this problem:

"Excessive formalization has proved to be organizationally dysfunctional, creating new layers of bureaucracy engaged in contract specification and monitoring, quality control, inspection, audit and review and diverting the energies of professional staff away from service and program delivery into a regime of form-filling, report writing and procedure following which is arguably even more extensive than that which existed during the former bureaucratic era."

The above statement gives a good starting point to reflect on the roots of NPM which has been inspired by methods used in manufacturing industry for improving quality. The public sector and other organizations which are conducting more abstract and creative work might not readily adhere to the criteria's needed to successfully become data driven, as the tasks are not easily measured, and hence a good feedback loop is hard to envisage. Below follows yet another problem of NPM:

"It might sound paradoxical, but stressed and de-motivated, 'unable' and/or 'unwilling' employees fit quite well into the ideological framework of NPM. Such aspects underline the necessity of more policies and procedures, of more systematic performance measurement and appraisal, of more monitoring and advising, of more 'leadership' and 'motivation' - for the whole arsenal of managerial concepts and methods." in [5].

The two data driven organizations from the former paper, Uber and Lyft, can ignore the problem above as their workers are not employed by them and hence they can skip the motivation part and fire them (or, rather, ban them from using their app) if they do not follow their minimum criteria.

We conclude this section with a summary: it exists a lot of driving forces for organizations to become data driven, efficiency being one of the more important ones. As we have noted by the anecdotal NPM-quotations above, not all tasks within an organization can be easily transform into a data driven tasks. We will elaborate more upon which criterions are needed and/or sufficient for a successful data driven tasks creation in later sections.

3 Interpretability of Inductive Models

Interpretability of inductive models has recently attracted more attention, there might be many reasons for this, but one of the most important reasons is that the (actual or planned) usage of inductive models start to become more common in organizations. For each organization this can pose new opportunities and raise new questions, as we have seen earlier when models or algorithms manage workers it can be seen as both cold and in-humane, but it can also be seen as an opportunity as all workers is treated equal.

In his [7] overview paper of the area of interpretability he identifies that there is not clear technical definition for interpretability at the moment and that different researchers do put emphasis on different properties and aspects related to interpretability, be it: trust, causality, transferability, fair and ethical decisions. Demand for interpretability can come from the users of the model or by regulatory agencies for example General Data Protection Regulation (GDPR) in the EU, which include, the right for citizens to contest automatic profiling from algorithms. This implies that European citizens have the right to know and understand algorithmic decisions, which in turn, means that an explanation must be provided for a particular case together with a prediction of a model.

It might then not come as a surprise that new methods for facilitating interpretability for particular predictions have been devised lately. These methods typically produce an explanation for a particular classification by introducing a local explanation model as a mediator between the instance and the global model. One such method is Local Interpretable Model-agnostic Explanations (LIME) [8]. One way LIME present its explanation is by showing the support in favor of and against the prediction.

One example of a method for facilitating interpretability is described in the paper [11], here the authors aim not only to explain why a certain prediction is given on a certain instance, but they also investigate how (with minimal effort) to change the prediction output via features which can be adjusted. This type of explanation could be very important and one that we often (as humans) take for granted. For example, let's imagine that you go to the doctor and get a diagnosis that you will have a heart attack within six months, then you are told to go home. This information is not sufficient to be helpful for you. After given such a diagnosis you want to receive advice on the changeable features, i.e. factor you can affect to avoid being in the class of patients that you were diagnosed as.

As mentioned already one of the reasons of creating interpretable models is for establishing trust in the models, here the assumption is that if the models are understood then trust will emerge. But is the interpretability of models necessary? If we have a model that behaves as an oracle in that it always makes the correct prediction, do we then need an explanation for the predictions? I would believe that most humans eventually would accept the models predictions without needing further explanation due to the oracles previous results in successful predictions.

4 Use-Case and Proposed Framework

Here we will describe the use-case we have studied and then present a data driven framework for supporting the use-case.

4.1 Use-Case

In this case study we will look at the situation at a European based globally present manufacturer, OEM, of commercial vehicles. We will in this use-case focus on the maintenance offer by the OEM and the demands on it. Traditionally manufacturers of trucks and buses have sold their vehicles to fleet owners and also offered maintenance contracts for these vehicles. This is still the case, but there are plans to provide transportation (tonnage/km) or bus (operation/hours) as a service. By using such a service, the fleet owner shifts into a fleet operator. In this new business model, the fleet owner has to pay a fee for the transportation service and is in return guaranteed an uptime percentage, i.e. the time when the transportation service can be utilized. Downtime, when the transport service is unavailable, can be divided into planned downtime, and un-planned downtime. Minimizing or eliminating the un-planned downtime, i.e. typically when a vehicle has broken down, is desirable. Planned downtime could for example be when maintenance is conducted on the vehicle, if this can be planned in a flexible way, it can often be aligned with a time when the fleet operator does not need the transport service.

Transport as a service puts high demands on knowing the actual health status of a particular vehicle to avoid un-planned downtime. With the aim of providing more uptime and only planned downtime, the OEMs maintenance program have in the last couple of years gone from a fixed interval (km and/or hours) based maintenance program to a data driven maintenance program. In the new program, the models for expressing the maintenance needs for particular components are both induced by machine learning algorithms and created manually by experts. While the maintenance models created by experts do not needed to be interpreted (at least not by the experts them self), there is a need to simulate the consequences of their manual crafted rules, to be able to validate them on a bigger population. The maintenance experts have a production tool which supports creation of expert rules and the functionality to simulate these rules on the current population before putting them into production.

The models produced by machine learning methods are typically validated by both their performance metrics and by manual inspection of the models. The latter demand has put restrictions on the type of models which can be utilized, typically they should be transparent, i.e. not black boxes, and not too complex. These demands for simple models put limitations on the performance of the machine learning methods. As we have discussed earlier the feedback loops in data driven organizations is essential for these types of organizations to work properly. At the OEM the main feedback loop was designed to quickly be able to respond to quality issues w.r.t. sub-series of components etc. But this feedback loop has not yet been modified to align with the new maintenance program.

The transport service offer that the OEM makes towards potential customers is typically made by experienced sales personnel who together with the potential customer investigate the needs of transport service provided by the OEM. If the customer already owns a vehicle from the OEM then the data from previous operations can be used, for specifying which new product that would suit the customer best and also estimate the cost for the transport service offered to the customer. If the customer changes utilization of the transport service, the customer and the OEM needs to re-negotiate the terms of the customer offer.

4.2 Proposed Framework

Given the use-case we will now present a suggestion for a framework that will support data driven ambition for the maintenance offer, the framework is shown in Fig. 2. The framework is initialized to the left in the figure with the alignment of product development with model development. Each model is created for a specific component and should be able to with a high certainty predict when the component needs maintenance to avoid failure. If the model predicts need for maintenance (maintenance decision) before failure, the model is working correct, but can waste resources, feedback of changed components needs to be analyzed to make sure that components are not changed prematurely. If the model fails to predict correctly and a failure of the component do happen, its cause should be analyzed (analysis). The outcome of analysis could be: the model is valid, there is a need to set off the alarm to alert humans that the models needs attention. This might trigger a number of actions, one being the need for analyzing the problematic cases together with the model to understand the problem and refine the model by these findings to cope with similar cases in the future. Having presented the general framework, we investigate how we can improve the OEMs service offer using it, the first step is to integrate the creation of a data driven models as a natural step in the product development process (left in the figure), this is not the case at the OEM now. There have been efforts in this direction for certain components by individual maintenance experts and keen development engineers working together, but this work needs to be a mandatory step in the product development process. Questions like: Who will have responsibility of the maintenance prediction model over time? Who will create data driven models? When should Machine learning or other modelling techniques be used? Needs to be answered.

My suggestion is that for each new product development effort at the OEM, create a team of maintenance experts, data science experts and the development engineers. The first task of the team is to assess the impact of component failure from different viewpoints, be it road safety, fulfillment of transport service, environmental aspects, etc. These viewpoints would then serve as input into how much time, effort and money the team can use for creating prediction models. The most expensive model being a physical model hand-crafted by senior engineers and should only be constructed for very critical components. Data driven models are cheaper to create but they rely on historical data of failures. If data driven models are chosen, a plan for collecting data from test, be it from rigs, to

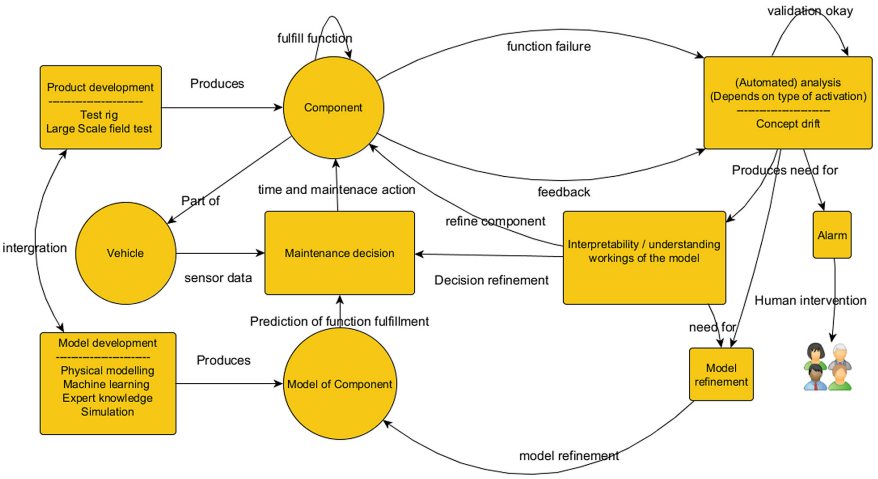

Fig. 2. A schematic view of the suggested framework

simulate usage, and later when the component is out in field test before release to customers, must be devised. Finally, if the model need not be of highest fidelity, the maintenance experts together with engineers will express their knowledge by formulating rules for component maintenance in an expert system. Interpretability of the data driven models created might show interesting patterns given the sensor inputs, i.e. engineers might gain insight to why certain types of failure occur. This is one of the major reasons of supporting interpretability at this stage. But this need not affect the performance of the model, instead model agnostic methods for interpretability should be preferred.

Once the predictive maintenance models have been put into production, they need to be evaluated and monitored in a smart way (the middle and right part of the figure). Typically, new models just put into production needs to be monitored more closely, compared to well-known models with known performance. But also these well-known models that seem to work well needs to be kept under surveillance. The question is how often and in what way to monitor the models, one obvious way is to create a feedback loop, one of the keys to becoming data driven. The feedback in this case should, ideally, contain a measure of how close to failure the component was and the prediction and confidence of the model. Here confidence is to be understood in the broad sense depending on the type of model we are dealing with. Given the type of component we are talking about, if it is replaced, as for example oil, belts etc. these can be analyzed to identify its remaining useful life, either directly by technicians or sent to experts. If the component involves interaction with the vehicle but do not render in objects which can be sent for analysis, one can use mediators or special instruments. These types of maintenance could for example be lubrication of parts, where a mechanic can do an ocular inspection and then report the assessment as feedback. In this case a picture could act as a mediator or we get a score could be given

by the mechanic. An example of special instrument is the instrument used to measure the health status of batteries or an oil quality test equipment.

The best feedback loops should (1) be non-intrusive, (2) require no extra work effort and (3) should be transparent, i.e. the motive of the feedback loop should be clear for the organization, one motivation for this is to avoid the frustration that was common with drivers from the study by the ride-sharing companies presented earlier. The first two criteria's can be motivated from our short overview of the problems with NPM. If there is need for much paperwork or other manual processes or there is no easy way of measuring the outcome, one should consider abandoning this feedback loop, as it will not probably be successful over time.

These criterions can be for example applied to the battery health instrument used by the mechanics in the OEMs workshops today. This special instrument gives an answer to the batteries health status, and the mechanic needs to write the results into a (digital) work order report. It scores fairly high on point 3 above, but low on the other two points. If we updated the special instrument to work in alignment with the work order report (the feedback loop), where the value is feed into the work order report once the measurement is done, it would the score high on all three points, which is a situation to strive for. Due to the diversity of components needing predictions, feedback loops will have to be created in many creative ways to be able to evaluate and monitor the predictive models. The feedback loops serve two major roles, one is to verify that the models are valid, and also to adjust and refine the model as we gather more data. Both of these tasks can be automated and decision makers can put thresholds on alarms when human attention is needed, for example if models no longer can be considered valid.

As the cost for realizing feedback loops varies depending on different components, establishing a good feedback policy would require a cost analysis together with the already mentioned analysis of criticalness of component and hence model. Once such a base policy has been established, it dictates how often feedback is requested for a given component. The next question is how to best select which cases to sample? Random sampling is not a bad idea, but a fraction of the sampling can also be directed towards cases where the model is uncertain. This sampling policy would be mostly beneficial when the model is new and exposure to new cases are more frequent and needed. In general interpretability of models is not of big importance, instead other measures of model performance are needed, be it accuracy or the like. In special cases when the models no longer is valid or if experts need to understand the models for other reason, it would be sufficient to use model agnostic techniques to interpret the models.

The feedback loop could also change the transport service offer to become more dynamic, so given how and under which circumstances the customer utilize the transport service, the price is adjusted in real time.

5 Conclusions

It this paper a use-case has been investigated for a company that is in the transportation manufacturing industry which is an industry in transformation from only selling hardware to selling different levels of services together with the hardware. The focus has been on the maintenance program, how the program can become data driven, and the implications both on the organization and on the business. A framework has been proposed for facilitating the data driven business model w.r.t. aspects such as: Need for interpretability of models; How to create the necessary feedback loops; Which type of predictive models to use and when. In our framework we also try to draw lessons from what's been done earlier, like in the presented ride-sharing case and examples from NPM to avoid pitfalls which a data driven organization are facing. This framework will hopefully serve as an inspirational starting point for other organizations that are facing similar business challenges as the OEM.

References

1. Anderson, C., Brown, C.E.: The functions and dysfunctions of hierarchy. Res. Organ. Behav. **30**, 55–89 (2010)
2. Diefenbach, T.: New public management in public sector organizations: the dark sides of managerialistic enlightenment. Public Adm. **87**(4), 892–909 (2009)
3. Hammer, M.: The process audit. Harvard Bus. Rev. **85**, 111–119, 122 (2007)
4. Hoggett, P.: New modes of control in the public service. Public Adm. **74**(1), 9–32 (1996)
5. Karp, T.: Unpacking the mysteries of change: mental modelling. J. Change Manage. **5**(1), 87–96 (2005)
6. Lee, M.K., Kusbit, D., Metsky, E., Dabbish, L.: Working with machines: the impact of algorithmic and data-driven management on human workers. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI 2015, pp. 1603–1612. ACM, New York (2015)
7. Lipton, Z.C.: The mythos of model interpretability. *CoRR,* abs/1606.03490 (2016)
8. Ribeiro, M.T., Singh, S., Guestrin, C.: "why should i trust you?": explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2016, pp. 1135–1144. ACM, New York (2016)
9. Shearer, C.: The CRISP-DM model: the new blueprint for data mining. J. Data Warehous. **5**(4), 13–22 (2000)
10. Srinivasan, V.: The Intelligent Enterprise in the Era of Big Data, 1st edn., Wiley (2017)
11. Tolomei, G., Silvestri, F., Haines, A., Lalmas, M.: Interpretable predictions of tree-based ensembles via actionable feature tweaking. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2017, pp. 465–474. ACM, New York (2017)