

Data-driven decision-making

Ahmad Saeed Khan-Örebro University

It is strategically significant for industrial operations to optimize decisions on a daily basis using data. Optimization techniques help to find optimal decisions by minimizing or maximizing an objective function subject to constraints. Optimization problem (OP) becomes complex to solve when the objective function and/or the constraints involve one or more unknown parameters. In this case, historical data can be used for the estimation of parameters by using coupled or decoupled approaches. However, historical data does not contain counterfactual outcomes which are essential for accurate OP solution, inference of counterfactuals from historical data is a daunting task due to presence of confoundedness and selection bias. Moreover, it is also important to make whole process of learning and OP formulation computationally feasible so that it can be applied to real-world use cases.

My PhD is focused on solving the above-mentioned problem by finding optimal decisions when historical data is of the form $\{x_i, y_i, z_i\}$, where x_i are features of the i^{th} sample, y_i is the outcome and z_i is the historically assigned decision. A blend of Machine Learning (ML) and Mathematical Programming (MP) provides an applicable foundation to solve OP under uncertainty based on historical data.

The rest of this essay talks about three areas of software engineering and shows how these areas are related to my PhD and how these can be utilized to effectively conduct my research.

Human factors

I am part of CoAIRob industrial school at Örebro University, main aim this school is to make intelligent systems and robots more effective by keeping the humans in the loop. I am working on system that would provide optimal decisions for future work based on historical data. In this research humans have very significant role to play majorly in two ways:

Formulation of objective function: Objective function and constraints are part and parcel of optimization problem. Objective function hinges on domain and specific problem in hand. Therefore, it is essential to develop objective function in collaboration with domain experts so that we solve actual and real-world problem instead of solving academic or general problem. Objective function formulation is first and foremost step of a decision support system if it goes wrong then whole exercise of problem solving would become futile.

Human evaluation of developed system:

System that are used to suggest decisions are crucial they affect the lives of masses directly, for instance, in medical setup decision produced by the system is a matter of life and death; similarly in financial settings it is a matter of prime importance. Therefore, it is paramount that developed system goes through the Stainer of human evaluation before it is deployed and start affecting the its audience. In my case, system is being built using historical data but humans that are working in domain possess plenty of experience and would have seen situations which are unknown to the system, that's why they are capable of judging the authenticity of suggested decisions.

To conclude, although we are heading towards the development automated processes everyday but still there exist areas where importance and necessity of human hand cannot be overlooked.

Security and privacy

Security is crucial in my research because:

1. Affecting the masse intruders might achieve ulterior motives by generating their desired decision from system. Therefore, it is important to make system highly secure
2. Might introduce samples in the online training datasets that can create undesired decisions.
3. On the other hand, I might use some real-world datasets collected by companies from real users. Therefore, it is essential for me to keep the identity of users anonymous to ensure the complete privacy of customers.

I need to keep both data and machine learning models secure by applying best practices of software engineering and using third party cloud services instead of securing with help of in-house capabilities. Privacy is major concern of my project because my project is utilizing the data which has been collected over the period from the customer. I should seek consent first and need to anonymize the data. I should also need to look whether data is biased against certain community and to ensure it is bias free.

In conclusion, security and privacy have important role to play in my research throughout the whole pipeline of my project.

Regulation and compliance

My project is being developed for H&M, therefore all rules and regulations exist here in Europe regarding data handling and privacy must be met. H&M is Sweden based company. It is bound to follow all rules and regulation that have been laid by GDPR. Values and principles that are more important to data-driven decision-making are fairness (my system should be bias free and diverse, it should not further increase gap between under and overrepresented communities) and transparency (all stakeholder should know why and how the certain decision has been made, if any complain is made then all stakeholder should be informed the underlying decision mechanism)

Data management

Sources

I am using data from two sources:

1. Online benchmark datasets for the estimation of ITE (individual treatment effect) which includes IHDP, Twins, Jobs and many other publicly available datasets. These datasets are owned by different governments and have been collected on various times as result of multiple programs.
2. Secondly, I would use real-world datasets collected by H&M group over the period as result of transactions by its customers around the world.

Sharing

I am employee of university (under the school of CoAIRob) but my industrial partner is H&M group. I have signed Nondisclosure Agreement with H&M group and I am bound to protect data and other confidential information conveyed to me by H&M group.

To conclude, Onus of benchmark dataset management does not really on me but when it comes to H&M real-world dataset I need to comply Nondisclosure Agreement (about ethics and code of conduct) signed between H&M and myself.

My thought:

I think machine learning and software engineering are two distinct paradigms. It is not like ML would eat software engineering. Both disciplines are sharing some commonalities and augmenting each other to flourish. Although the way AI\ML solution development is different, but it does not mean that it is beyond the realm of software engineering. If we observe meticulously, duties of software engineering team in traditional development process overlap with the duties of machine learning engineering team. I believe both paradigms will continue to progress without losing their identities.