SYSEN 6000: Foundations of Complex Systems

Stock to Flow, System Boundaries & Principles

Nick Kunz [NetID: nhk37] nhk37@cornell.edu

September 23, 2022

Stock to Flow: Synthetic Data & Decision Making

This analysis broadly illustrates the stock and flow relationships between system level boundaries and its principles thereof. More specifically, it addresses the general application of synthetic data in data-driven decision making and is meant to serve as a basis for future research in that regard.
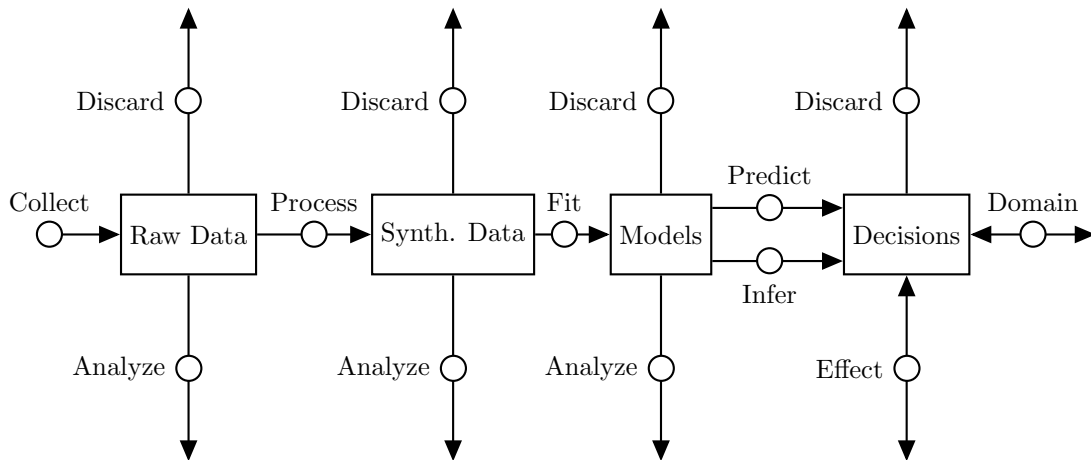

Fig. 1: Stock to Flow - Synthetic Data in Data-Driven Decision Making.

System Boundaries & Principles

A System of Boundaries

The following description of synthetic data in data-driven decision making outlines the following stocks and flows in reference to *"Fig. 1: Stock to Flow of Synthetic Data into Decision Making"* on Pg. 1. There are 4 stocks contained within the system of interest, they are: 1) *"Raw Data"*, 2) *"Synthetic Data"*, 3) *"Models"*, 4) *"Decisions"*. The contextual bounds of these stocks were specified in order to explain the broader system boundary.

For each stock, exist outflows, where any unit of a given stock can be removed from the system. This can be seen in the outflows, *"Discard"*. Inversely, each stock also has an outflow, *"Analyze"*, where each unit of stock can be analyzed directly, with the exception of the *"Decision"* stock, where the bi-flows, *"Domain"* and *"Effect"*, can interact with a distinctly different system beyond the scope of this one and is discussed later.

The primary inflow begins at data collection, *"Collect"*. This was important to define as a broad system boundary, as our system of interest does not address how and for what reason the data is collected. The raw data is assumed to be collected by the broader system of interest independently from how it was generated or composed. The system of interest is only concerned with the capacity in which better decisions can be made from it.

After the raw data is stored within the stock, *"Raw Data"*, if it is neither discarded nor analyzed directly, it is then processed into synthetic data and stored within the stock, *"Synth. Data"*. The outflow of raw data to the inflow of synthetic data, through *"Process"*, is the moment where the system of interest becomes distinctly different from a typical setting, where raw data is likely to be pre-processed, but not deliberately modified to reflect direct shifts in its distribution.

Moving from the synthetic data to models, if the stock, *"Synth. Data"* is neither discarded nor analyzed directly, it outflows to the inflow of the stock, *"Models"*. This occurs with the outflow, *"Fit"*. This is typical for settings where models are useful. Models are considered fit to the synthetic data received from the stock *"Synth. Data"* in cases where they are used for prediction, inference, or both. In this sense, model *fitting* can also be thought of as model *training*, depending on which nomenclature is most appropriate.

The penultimate step between models and decisions are two independent outflows, *"Predict"* and *"Infer"*. These become applicable if the models in the stock, *"Models"* are neither discarded nor analyzed directly. Making a distinction between the two independent outflows is important because it denotes the motivation for synthetic data in data-driven decision making in the stock, *"Decisions"*. In the outflow *"Predict*, the interest is focused on predictive capability at the cost of explainability. In the outflow *"Infer"*, the interest is fundamentally explainability, as a way to inform decisions.

The final steps beyond the stock, *"Decisions"*, are two independent bi-flows, *"Effect"* and *"Domain"*. The bi-flow *"Effect"* describes the behavior that the stock, *"Decisions"* has on real-world outside of the system of interest. The bi-flow *"Domain"* describes the domain knowledge that is utilized in data-driven decision making, which also originates beyond the scope of the system of interest.

The bi-flows *"Decisions"* and *"Effect"* were important to define as broader system boundaries, as our system of interest does not address how decisions effect the real-world beyond the system of interest, nor does it account for the origination of domain knowledge. Rather, it assumes that synthetic data in data-driven decision making includes knowledge of a particular domain outside of the system of interest, as well as how real-world outcomes influence decision making.

A System of Principles

The underlying principles exhibited by the synthetic data in data-driven decision making are most explicitly highlighted by 4 of the 12 system principles mentioned in *"Principles of Systems Science"* by Mobus and Kalton (1). The subset of 4 system principles explicitly captured in this analysis are: 1) *"Systemness"*, 2) *"Systems Encode Knowledge and Receive and Send Information"*, 3) *"Systems Can Contain Models of Other System"*, and 4) *"Systems Can Be Improved"*.

The system of interest is generally consistent with the *"Systemness"* quality described by Mobus and Kalton. Because the system of data-driven decision making with synthetic data heavily relies on software systems, and systems of the world, both captured through data, and interpreted through the mind, it can be reasonably assumed that the system of interest can be characterized as *"Systemness"*, as it inherently requires an intersection of system aspects including ontology, epistemology, mathematics, symbolic language, and affect.

In addition, the system of interest is generally consistent with the *"Systems Encode Knowledge and Receive and Send Information"* principle. This is most evident in the bi-flow, *"Domain"*, where the stock of decisions is influenced by the inflow and outflow of external domain knowledge. These decision, also effect future domain knowledge, as well as the real-world effect. Furthermore, at every step of the data-driven decision making process, the system naturally encodes knowledge through inductive bias, which begins even before the data is collected and is especially pronounced with in the inclusion of synthetic data.

It is also evident that the broader system of interest is generally consistent with the *"Systems Can Contain Models of Other System"* principle. Much like the *"Systems Encode Knowledge and Receive and Send Information"* principle, the *"Systems Can Contain Models of Other System"* principle emerges at every step of the data-driven decision making process. Each flow between each stocks, contains a more detailed system at a deeper level of analysis. In other words, we have a nested hierarchy of systems modeled underneath the broader system boundary.

Finally, the system of interest is consistent with the *"Systems Can Be Improved"* principle. No matter how clever the existing solutions for data-driven decision making with synthetic data become, they can always be improved. As previously mentioned, the system of interest heavily relies on software. Practically, the value of software development has been its maintenance and incremental improvement. If this value is to continue into the future, then it is reasonable to assume that, not only can the system of interest improve, but its value increases because of it.

Final Thoughts

The system of synthetic data in data-driven decision making is bounded to a typical path of stocks and flows that naturally emerge in application. The bounds of the system are defined by what is most useful and immediate for its stated goals and is generally consistent with the set of system principles described by Mobus and Kalton in *"Principles of Systems Science"* (1). This analysis has outlined 4 of the most obvious consistencies, but it is likely that the system of interest contains all 12 of the qualities mentioned. In future analyses, it might be useful to detail all 12 of the principles, expand the details of the lower level systems, provide a practical demonstration of the stock and flow relationships, or perhaps some combination of the three.

References

[1] G. E. Mobus and M. C. Kalton, *Principles of Systems Science.* Springer, 2014.