AGPhil 3: Symposium: Epistemology of Big Data in Physics I

Zeit: Donnerstag 13:30–15:30 Raum: GW2 B2900

Hauptvortrag AGPhil 3.1 Do 13:30 GW2 B2900

Data-driven hypothesis generation using deep neural nets — •Balázs Kegl — CNRS / Université Paris-Saclay

Generating and testing a large number of low-probability hypotheses in certain scientific fields lead to the so called p value controversy. From the point of view of hard sciences this seems as an abnormal misuse of the scientific method. In the first part of the talk I will argue that the scientific method, as it is understood today, does not prevent these aberrations. In tomorrow's world where computational tools can generate scientific hypotheses automatically, fixing this issue is of uttermost importance. Solving this problem will require putting hypothesis generation back into the center of the scientific method.

The goal of computational creativity is to design methods that can generate valuable novelty. One major debate within this community is whether generation is mostly random (only the evaluation process has a strong notion of value of novelty), or we should include knowledge already in the generative process. I will show how this issue is related to the p value controversy and automatic hypothesis generation. I will present a constructive framework in which data- and knowledge-driven novelty generation can be studied and evaluated. I will finish the talk by showing some of our latest results using deep neural nets as the knowledge representation and novelty generation engine.

Hauptvortrag AGPhil 3.2 Do 14:15 GW2 B2900

How can we learn useful things from big data? Data mining from the perspective of Meno's problem — •Claus Beisbart — University of Bern, Switzerland

In modern physics, many data sets arise not because there is theoretical motivation to study a set of variables, but rather because new instruments allow for the speedy accumulation of huge sets of measurements. An important challenge then is to make scientific use of the data. As L. Floridi puts it, the challenge is to find small patterns in big data. The aim of this talk is to understand how methods of data mining may meet this challenge.

I approach this topic from the perspective of a puzzle presented in Plato's "Meno". There, it is argued that we cannot search for something yet unknown (nor investigate something yet unknown). For to claim success, we would have to have a criterion of success, and such a criterion may only be given if we knew what we are searching for, which we do not. Whereas the paradox can be resolved in a rather trivial way for many searches, it has more plausibility in the context of big data, because scientists are there looking for something they don't have any clue about.

My philosophical project thus is to explain how data mining may produce new knowledge despite the paradoxical conclusion from "Meno". I do so by presenting a case study from astrophysics and by analyzing representative examples of methods of data mining. My focus is on the aims of, assumptions behind, the methods.

AGPhil 3.3 Do 15:00 GW2 B2900

Collaborative scientific practice, epistemic dependence and opacity: the case of space telescope data processing — •Jebeile Julie — Université catholique de Louvain, Belgium

A great part of scientific knowledge is the outcome of a collective enterprise supported by technologies, mainly instruments and computers. Astrophysical images, in particular, are today built from data whose measure and digital processing involve competent teams of astrophysicists, including instrumentalists, experts in computer programming and specialists in data analysis, as well as technologies, including telescopes and computers, at different steps of the processes. Here scientific knowledge crucially depends on the trust agents place on their co-workers and on the required technologies.

In such a scientific context, an agent cannot trust one another by merely appealing to her intellectual authority, contrary to what epistemologists sometimes suggest for more ordinary epistemic situations. The agent rather must have sufficient evidence for the trustworthiness of her colleague's inputs. However, based on the case study of Herschel space telescope data processing, I argue that such evidence is sometimes not accessible to her for several reasons on which I elaborate. In this case, she more or less opaquely depends upon her collaborator epistemically. Yet opaque epistemic dependence is certainly not desirable in the process of producing scientific knowledge. As I show, the same holds for the use of instruments and computers. The scientists who actually rely on instruments and computers do not all have access to evidence for the trustworthiness of the outputs.