Big data hype hasn't led to tangible results in the social sciences, expert says

March 14 2018

Data science can help to answer research questions in the social sciences. Credit: Colourbox

We've already got used to the fact that our digital footprint is continuously recorded, stored somewhere and evaluated. This has revolutionized the advertising industry, and companies like Uber and Amazon are using our data to be even more efficient. Discussion of how digitization is affecting our lives, however, is often limited to speculations about what Google or Facebook might do with this data.

Despite the great progress in basic research, such as speech recognition and image processing, success stories of existing big [data](#) applications in the social sciences are scarce. As early as 2014, big data plummeted from the "Peak of Inflated Expectations" to the "Trough of Disillusionment" phase in the [Gartner Hype Cycle](#). In the basic sciences, the focus is on the technical prerequisites for efficiently recording and storing large quantities of data and automatically processing them. Artificial intelligence methods such as machine learning have great potential here. Only the social sciences have so far benefited little from this, and even seem to be losing ground to other disciplines. I notice that instead of drawing benefit from the flood of data for their empirical research, [social scientists](#) are often overwhelmed by the opportunities that arise.

The void opening up is filled by other scientific disciplines – engineers collecting sensor data on individual mobility, for example, and computer scientists extracting statistical models from such data. This data-driven approach to [social phenomena](#) is now often referred to as computational [social science](#). Recently, one had the illusion that the classical approach of the social sciences – hypothesize, model, test – would become obsolete; instead, a new form of social [science](#) would emerge in which theory is replaced by machine learning of social "laws" from the data.

Data science can indeed help to answer research questions in the social sciences; but it cannot develop such questions by itself. The "discovery" of statistical correlations cannot replace the scientific clarification of causal effects. For in the social sciences, questions are not only about "what", but also about "why". Social scientists are therefore indispensable in order to make computational science a social one.

What's required are novel models of [social interaction](#) that are expressly developed bearing in mind their calibration and validation against large, previously unavailable quantities of data. This calls for a new

methodological expertise, and it is up to the universities to teach it. At the Chair of Systems Design, we have taken the challenge by developing courses on the theory of complex networks, agent-based modelling of [social systems](#) and statistical analysis of social data.

The opposite also holds true: the engineering sciences can benefit from the social sciences. Technical systems today are dependent on the social dimension – their users. It's not feasible to design a smart energy supply or a shared platform for software development without considering human behaviour and social relations –- and this is exactly where the core skills in social sciences lie. An interdisciplinary training of engineers and computer scientists is called for. Right now, while the foundations of computational social science are still being laid, We have the chance to work together across borders. I'm convinced that this will determine the success of the disciplines – on both sides.

Provided by ETH Zurich