

Big Data ? Data + Questions + Algorithmic Knowledge + Commitment + Collaboration - Part II

Autor : Reinhard Riedl

Datum : 24. Februar 2017



Many big data activities fail on establishing necessary preconditions for success. For companies the main challenge in applying big data is to ask the right questions. In science the main challenges are the access to data and the collaboration across disciplines.

[Read Part I of this article](#)

In search for skills and sharing

Having enough data and proper questions is good, and for many companies it is good enough, but not for all and sure not for more advanced scientific research. In many situations applicants of big data need deep algorithmic knowledge, lots of practical knowhow, and advanced tools. At least, they should understand how they can use techniques from data science to work with data of mixed quality, how they can eliminate bias from data, and how they can focus on the relevant dimensions in high dimensional data spaces. They should know in which cases under which conditions there are tools and experts to help them with these and other challenges. I.e., they should be clearly aware of the potential (and the limitations) of the contemporary algorithmic

machinery and include it into their projects whenever needed, possibly through external contributors.

Because the wide range of necessary expertise, in practical settings applied data science often relies on the results of many people/teams. For good reason it is preferably performed jointly by a multidisciplinary team (of teams) combining many different competencies and cultures. This brings us to the potentially biggest challenge of big data, the dramatically growing importance of cross-disciplinary collaboration. In the above example of big data in healthcare research the team (of teams) may include more than half a dozen disciplines from clinical specialists to pure mathematicians. Clinical doctors may thus end up counting Brownian motion paths – or at least they will have to judge on the relevance of the results of such a counting. In such settings, it is critical that the cross-disciplinary collaboration works.

Cross-disciplinary collaboration is always tough. There is ample evidence that collaboration fails if it relies on interfaces. Instead it is necessary that the flow of results among the various disciplinary experts/teams is nurtured by domains of shared knowledge. One very basic technique to establish these domains is to exploit a series of boundary objects.

Anything goes: coaches and change

All the above does not sound like good times for consultants. Managers have to ask their questions themselves, advanced mathematics is needed, and experts have to establish/build domains of shared knowledge – what are then consultants good for?? Well they are needed – urgently needed – as coaches for individuals in big data ventures and as life cycle managers for the boundary objects. Some of them may even ascend to team (of teams) coaches if they are willing to take responsibility for the outcome of big data ventures. We do need team coaches that combine an integrative understanding of many disciplines in renaissance style with contemporary high performance leadership qualities as we know them from sports teams, or music ensembles, respectively.

Beyond buzzwords and sales speeches big data is big mess that needs artistic inspiration. It is like going down the bullshit creek without a paddle in order to compose a piece of music. This image does not make sense to you? Well, the same holds for many big data ventures – but in some of them the image does make sense. When big data becomes a tool to save a human life, the discussion about hype or not loses its meaning. It then becomes clear that we need both(!) knowledge and knowhow and that an excellent scientific education is very helpful for orientation purposes.

Big data truly is anything goes in the sense of Paul Feyerabend. That means, that projects have to fit their methods to the case and probably adapt during runtime. Of course, scientifically speaking, that is forbidden in many disciplines – and it is exactly this “crime” against scientific rules which provides the true thrill of big data projects. The greatest of them courageously leave established disciplinary grounds to discover new land. However, if they move too far away from known grounds, they may also drown in the sea of their own, data generated fictions. The list of childish confessions of faith in data science is already a long one indeed.

Summary

There all kinds of big data projects pursuing very different tasks – from very trivial to absolutely impossible ones. But for the more challenging of them, the inequality in the title holds: Big Data ? Data + Questions + Algorithmic Knowledge + Commitment + Collaboration. Thereby “?” should be read as “is more than just”, the difference being a great idea, good luck, and further ingredients (like a broad scientific education). Great ideas and luck cannot be enforced. The better you know a domain, the more likely is that you understand meaning and relevance of a new idea (and thus can exploit it), but the less likely is that you have one that is really new. So a proper mix of expertise and talents and a true commitment to collaboration remain the bottom line of big success with big data.