

Learning to analyze big data: some personal experiences

林金龍
國立東華大學財務金融系

Abstract

The term *big data* has become increasingly prevalent, and it appears that economists need at least to know something about it. Yet, as big data involves massive data collection, processing, cleaning, feature extraction, analysis and presentation, it demands deep knowledge about web crawling, processing tools and data analytics, not to mention appropriate statistical methods. As a beginner in this field, I have spent a great deal of efforts and time to learn about mastering big data. In this talk, I would like to pass on my learning experience. One can easily get drowned in the big data ocean as there are so many materials surfing the net, including news, blogs, slides, video clips, tutorials, codes, conference and journal papers. Thus, I shall give the talk in a question-answer format. Hopefully, by answering the frequently asked questions, I could draw a clear picture about big data and shed some lights on learning to crunch it. Here are the questions.

1. What is big data?
2. Why is big data so hot?
3. What are the similarities and differences between data mining and statistical analysis?
4. How is big data different from small data?
5. Should an econometrician learn about data mining or can one survive without touching big data?
6. How to learn about data mining?
7. What are the popular techniques for data mining?
8. What softwares to use for big data?
9. What are the essential skills for big data analytics?
10. Care to recommend a few good books and websites for the beginners?
11. What should I do should I decide to become a data scientist?
12. Could you give me some real examples with source codes?

To my mind, the best way to learn about big data is to analyze one. I shall cover several real examples in this talk.