

Source: <http://www.opengardensblog.futuretext.com/archives/2013/10/my-course-on-big-data-algorithms-for-smart-cities-at-city-science-upm-madrid.html>

By [ajit.jaokar](#) at [futuretext.com](#)

My course on Big Data algorithms for Smart cities at City science – UPM (Madrid)

October 5, 2013 By [ajit](#) [Leave a Comment](#)



Background

I have blogged before about the need for [algorithm transparency](#) for [Big Data algorithms for Smart cities](#) . The same sentiment is expressed in [Rage Against the Algorithms – How can we know the biases of a piece of software?](#) [By reverse engineering it, of course](#)

After a year or so, I have made some progress on the idea of Big Data algorithms for Smart cities and I will try and elaborate here in this longish blog post which you can download also as a pdf. In addition to my Oxford university course on Big Data for Telecoms, from Jan 2014 onwards I am pleased to be also teaching about Big Data Algorithms for Smart Cities.

This also includes IOT, Mobile and M2M data.

At the newly launched [City Sciences](#) program at UPM – [Technical University of Madrid – Universidad Politécnica de Madrid](#), I will be teaching about applying Big Data algorithms (specifically Mahout, Real time algorithms like Twitter storm, Predictive algorithms and Machine learning algorithms) to Telecoms, IOT and Smart cities

I am excited about this and always wanted to do this!

Also, I don't see many other places where this is being done .. so it's truly pioneering. Spain is a hotbed of Smart city and mobile activity especially with initiatives like [Smart Santander](#) and the [GSMA Connected Living initiatives](#)

One of the reasons for this blog post is to reach out to companies and other researchers who are working in this space (ex [IBM Smarter planet](#), [SAP](#), GE([Industrial Internet](#))) are all doing some interesting work in this space – as are research institutes like [fraunhofer FOKUS](#)). I am already doing some interesting work in this space especially at Liverpool Smart cities projects – [Connected Liverpool](#) - so we are already looking at real world applications

Another way to look at it is to think of the role of a Data Scientist for a city - The Harvard business review says that [the role of the Data scientist will be one of the hottest roles going forward](#)

Here are some ideas about my thinking

Note that this is not the actual syllabus – it shows more my thought process

Approach

My approach involves applying insights from one domain(Big Data algorithms) to data from Smart cities, Mobile and IOT.

So, we start with the maths – for example

[Differential calculus](#),

[Discrete maths](#),

[Probability theory](#),

[Linear algebra](#)

and then techniques such as

[Decision trees](#),

[nearest neighbour](#),

[unsupervised learning](#),

[Probabilistic modelling pdf](#),

[Bayesian learning](#)

[Predictive analysis techniques \(Predicting the future, What is predictive analytics – Part 1, Predicting the future, What is predictive analytics – Part 2\)](#),

[Machine learning algorithms](#)

[Real time algorithms like Twitter storm](#)

[Apache Mahout](#) etc

we then apply these to optimization problems based on data streams from Smart city verticals(like transportation), IOT, Mobile data and Open Data streams all within the context of the [R programming language](#) – albeit there is some great work on Python as well ex [scikit learn](#)

Why now?

Both IOT and Open Data are maturing .. many new initiatives will make IOT data increasingly common. Apart from mobile phones, apps and sensors – we also have initiatives like [alljoyn](#), [IFTTT](#) and [webinos](#) for IOT and Operators like Telefonica using [Open Data in innovative ways in partnership with the Open Data institute](#)

So, soon we will be presented with an abundance of Data. How to optimise it to get real insights will be the next challenge. Hence the algorithms.

This also brings us to Data. I was trying to find a taxonomy of mobile data. The closest I came to was this paper. Although from 2007, the principles still apply [Towards a Taxonomy of mobile applications\(pdf\)](#)

Mobile data streams

Candidate dimensions for a mobile taxonomy

Temporal dimension. (Synchronous: user and application interact in real time, Asynchronous: user and application interact in non-real time)

Communication dimension. (Informational, Reporting, Interactional)

Transaction dimension: (Transactional, non-transactional)

Public dimension: (Public, Private)

Multiplicity (or participation) dimension: (Individual, Group)

Location dimension: Some mobile applications provided customized information or functionality based on the users

location, whereas other applications do not depend on where the user is located.

The identity dimension relates to whether the identity of the user is used to modify the application based on the user's identity.

Categorization of Sample Mobile Applications

Purchasing location-based contents (local information, routing, etc.):

Mobile inventory management for a company:

Product location and tracking for individuals (e.g., searching for a certain plasma TV in a given city):

Mobile auctions:

Mobile games:

Mobile financial services (mobile banking):

Mobile advertisement (both user-specific and location-specific):

Mobile entertainment services (stored contents-on-demand, live events):

Mobile personal services (mobile dating):

Mobile distance education (synchronous and asynchronous versions):

Mobile product recommendation systems:

Wireless patient monitoring:

Mobile telemedicine:

So, potentially, all these applications (and many more from apps) could provide mobile data. We also need a taxonomy of city data

A taxonomy of City Data

Domains like Transportation will be early providers of City data – but in the blog

[Big data for Smart cities – How do we go from Open Data to Big Data for Smart cities](#) – I listed many more

Environmental data (particulate matter, CO₂, pollen)

Markets (weekly, flea, Christmas markets)

events (festivals, concerts, long night of ..., sports events)

Disposal (appointment in my street, recycling centers, container sites, hazardous waste)

infrastructure (cycle paths, toilets, mailboxes, ATMs, telephones)

Traffic (construction sites, traffic jams, road closures)

transport (delays, cancellations, special trips)

opening times (libraries, museums, exhibitions)

Management (Forms, responsibilities, authorities, opening times)

consumer advice, debt counselling

Family (parental allowance, day nurseries, kindergartens)

Education (schools, community colleges, colleges and universities)

Housing (housing benefit, rent prices, real estate, land prices)

health (hospitals, pharmacies, emergency services, specialist counselling services, Blood donation)

Pets (veterinarians, animal shelter, animal care)

Control (bathing, food, restaurants, prices)

Legal (laws, regulations, guidance, arbitrator, evaluator)

Police Online (current events, investigation, crime Atlas)

City Planning (zoning, construction, transport, airports)

Population (number, regional distribution, demographics, purchasing power,

Employment / unemployment, children)

And ofcourse wearable mobile data technology could create its own data streams

What makes a city Smart?

How do we bring this all together?

The ex Chinese Premier Wen Jiabo once said “**Internet + Internet of things = Wisdom of the earth**”

Indeed the Internet of Things revolution promises to transform many domains ..

As the term Internet of Things implies (IOT) – IOT is about Smart objects

For an object (say a chair) to be ‘smart’ it must have three things

- An Identity (to be uniquely identifiable – via IPv6)
- A communication mechanism(i.e. a radio) and
- A set of sensors / actuators

For example – the chair may have a pressure sensor indicating that it is occupied

Now, if it is able to know who is sitting – it could co-relate more data by connecting to the person’s profile

If it is in a cafe, whole new data sets can be co-related (about the venue, about who else is there etc)

Thus, IOT is all about Data ..

By 2020, we are expected to have 50 billion connected devices

To put in context:

The first commercial citywide cellular network was launched in Japan by NTT in 1979

The milestone of **1 billion** mobile phone connections was reached in 2002

The **2 billion** mobile phone connections milestone was reached in 2005

The **3 billion** mobile phone connections milestone was reached in 2007

The **4 billion** mobile phone connections milestone was reached in February 2009.

So, 50 billion by 2020 is a large number

Smart cities can be seen as an application domain of IOT

In 2008, for the first time in history, more than half of the world's population will be living in towns and cities. By 2030 this number will swell to almost 5 billion, with urban growth concentrated in Africa and Asia with many mega-cities(10 million + inhabitants). By 2050, 70% of humanity will live in cities.

That's a profound change and will lead to a different management approach than what is possible today

Also, economic wealth of a nation could be seen as – Energy + Entrepreneurship + Connectivity
(sensor level + network level + application level)

Hence, if IOT is seen as a part of a network, then it is a core component of GDP.

So, what makes a city 'smart'?

Building upon the previous discussion, my view is a Smart city is a city that behaves like the Internet i.e. is a platform/enabler for its citizens. Thus, the citizens make the city 'smart' by adding knowledge, value, data etc. This is a part of a wider socio economic trend to go from 'mass production' to 'smaller individualized services' – ex in music, in urban farming, in the Bristol pound, in local sourcing of food etc.

Holy grail – improved services

In conclusion, the payoff for a city is improved services. This is already happening for instance in a far of place like Abidjan ([AllAboard: a system for exploring mobility and optimizing transport in developing countries using cellphone data](#)) and in [healthcare](#) and we are seeing many new forms of radios like [Cell dot from Ericsson](#) and [Internet connected super highways using white space](#)

We could thus see a new value chain of sensor – Data – Algorithms – visualization

If you are a Vendor – company –researcher working in this space – happy to discuss solutions, joint papers etc. Pls contact me at [ajit.jaokar at futuretext.com](mailto:ajit.jaokar@futuretext.com)

Image – shutterstock