DOI: 10.26314/GARR-Conf17-proceedings-27

Motivating carsharing services open-data mandatory APIs

Andrea Trentini, Federico Losacco

Dipartimento di Informatica – Università degli Studi di Milano

Abstract. Modern countries car traffic is nowadays fought by regulations: tolls, dedicated/narrow lanes, low speed limits, reduced parking availability, etc. Some help can come from carsharing, i.e., pools of shared vehicles to be rented for short periods of time. The authors scraped carsharing websites for a couple of years, uniformed data and then queried and graphed the dataset, a summary of results is presented. Carsharing vendors should publish - openly, mandatorily, through standardized APIs - the state of their pool because this data can be used to analyse the overall traffic behaviour in town and study impacts, effectiveness and costs.

Keywords. Application Programming Interface, urban congestion, open data, public accountancy.

Introduction

Air pollution became a problem for human beings with the industrial revolution (Seinfeld and Pandis, 2012; Lave and Seskin, 2013; Steinle et al., 2013). Afterwards many governments begun legislating (US-EPA, 2013) to reduce industrial emissions. Then, air pollution slowly began to decrease as new generations of technologies replaced older ones (Figure 1).

Fig. 1
Historical trend of SO2
(source: ARPA Lombardia)

While pollution is being reduced, many governments and administrations are now addressing the "congestion factor" with regulations (Vanderbilt, 2009) to lower private traffic. A few years ago "carsharing" was introduced with the rationale that: 1) parking space is saved since a single parked vehicle serves a lot of users, not a single one (of a private

Conferenza GARR 2017 - The data way to Science - Selected Papers
Venezia, 15-17 novembre 2017 - ISBN: 978-88-905077-7-9 © 2018 Associazione Consortium GARR

car); 2) because of the higher (>private car, <taxicab) costs per use, many carsharing users, would use it wisely (maybe mixing it with standard public transport), reducing the overall "car mileage load" on the city; 3) it may further lower pollution (Firnkorn and Müller, 2011). A review on carsharing studies can be read in (Jorge and Correia, 2013).

GPSs and smartphones track down available cars and lead users to precise parking locations. Detailed location data are, of course, stored and secured into vendors' systems but some data are online and can be used to analyse the overall traffic behaviour in town. Current online data is referred to the real-time situation only and it is often behind "web-stacles" (Trentini, 2014). Given the usefulness of this data, the authors propose the initiation of a standardization and "open-data-ization" process. I.e., governments should force vendors to publish data (both real-time and historical) through public, documented and well-defined APIs (Application Programming Interfaces).

1. The quest for data (web-scraping)

In Milan (ITALY) there are five carsharing vendors: 1) Share'NGo (http://www.sharengo.it); 2) Twist-Car (http://twistcar.it, discontinued 17-Nov-2015); 3) Enjoy (http://enjoy.eni.com); 4) Car2Go (http://www.car2go.com); 5) GuidaMi (http://www.guidami.net). Some vendors do publish data about car positions and availability which the authors using wget in shell scripts or small python programs periodically run on a server.

2. Data analysis

The following analyses are based on data between September 2015 and March 2016, they show an evident daily and weekly patterns, i.e.:

1) every day is "cyclical", e.g., night differs from day and office hours are peaks;
2) weekly effects: a) Mon-Fri days are similar to one another; b) Saturday and Sunday are similar; c) Saturday and Sunday differs from Mon-Fri workdays;
3) there are, of course, random fluctuations.

Data were divided into weeks and for every week Q-Q plots (Quantile-Quantile, to check if samples are statistically similar) were created. A Q-Q plot for every day pair (Mon-Tue, Mon-Wed, Mon-Thu, etc.) in every week was generated, originating (7 on 2) = 21 Q-Q plots per week (examples in Figures 3 and 5). This procedure statistically confirmed the evident (and expected) pattern difference between Mon-Fri and Sat-Sun. Abbreviations: H = Holiday; NH = Non Holiday.

2.1 Available cars

The total number of free cars at a given time t is the sum of parked cars. When this value is low it means that many cars are in use. Figure 2 shows the usage for two typical days (from 00:00 to 23:59), one weekday and one weekend day. The "late Saturday" (Sunday early morning) usage is evident in the right graph. Other notable general remarks are: 1) night-time (between 2AM and 7AM) is a peak of free cars; 2) the usage peak (least number of free cars) is between 6PM and 9PM; 3) morning (between 8AM and noon) usage is lesser than afternoon (noon to 7PM) usage; 4) night-time peak sports a shorter timespan (between 6AM and 8AM) w.r.t. the Monday-to-Friday night-time peaks; 5) there is a usage peak between midnight and 2AM, often more substantial than the afternoon peak. In Figure 3 the Q-Q plots of two day pairs taken as example: 2015-10-05 against 2015-09-29 (Tue against Mon), near the y = x line → statistically similar; 2015-11-22 against 2015-11-17 (Sun against Tue), far from the y = x line → statistically different. I.e., (as expected!) users' behaviour on Sunday is different from a weekday.

Fig. 2
Free cars average: typical weekday (Wed) vs. weekend (Sun)

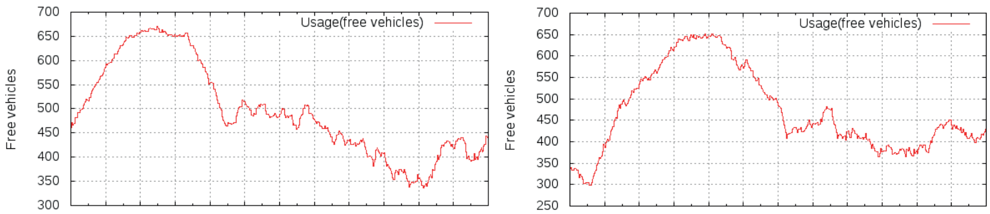
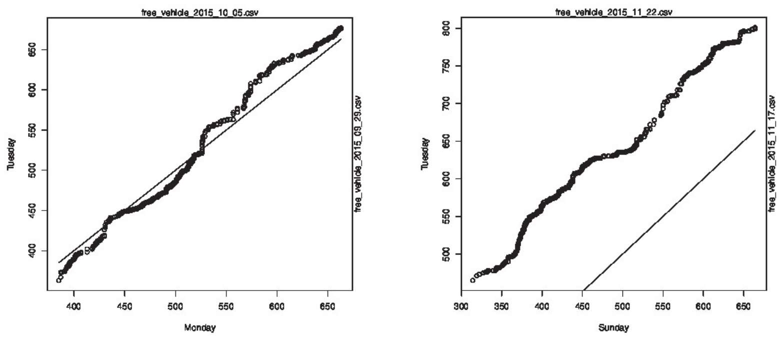


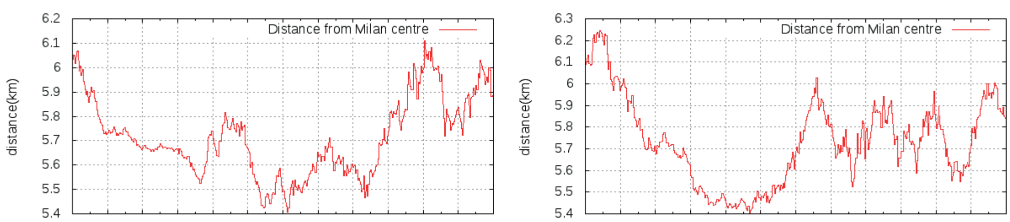
Fig. 3
Usage Q-Q plots: typical NH-NH vs. H-NH



2.2 Overall (average) distance from city centre

Notable remarks in Figure 4 ("Mon to Fri" left, "Weekend" right, they represent typical days from the dataset, X-axis is from 00:00 to 23:59) are: 1) the "lung" effect: average distance decreases during the day (8AM to 6PM) and increases during the rest of the day, i.e., people move into the city during daytime and move out of the city otherwise; 2) there is a concentration peak between 11AM and 2PM, i.e., during lunchtime users are closer to the centre; 3) between 8AM and 11AM there is a sudden out-movement of people closely followed by an in-movement, i.e., many users move, but not exactly at the same time; In Figure 5 the Q-Q plots of two day pairs taken as example: 2015-10-22 against 2015-10-21 (Wed against Thu). Again, users' behaviour on Sunday is different from a normal weekday.

Fig. 4
Average distances: typical weekday (Wed) vs. weekend (Sun)



It is also interesting to show the average distance of vehicles "per vendor" (Figure 6): 1) Enjoy is always more distant than the others (Car2go and Sharengo); 2) Sharengo (electrical only) was always the closest; 3) there has been a "getting farer" trend in Sharengo cars between 2015 and 2016.

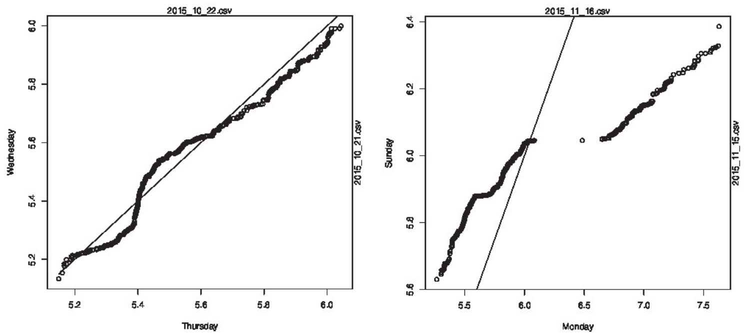


Fig. 5
Distances Q-Q plots: typical NH-NH vs. H-NH

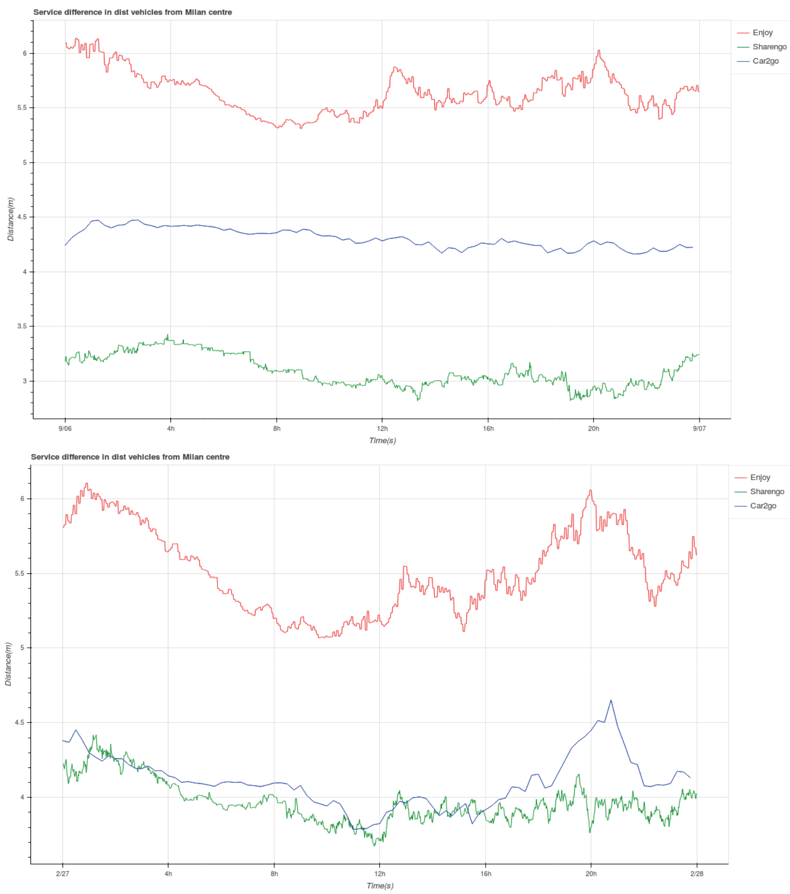


Fig. 6
Difference in average distances "per vendor" (top: 2015/09/06, bottom: 2016/02/27)

2.3 Movements

A "movement" is simply a vector (it can be plotted) between a car "catch" and a "release". The database of movements contains the records: current fuel, delta fuel, delta km, delta time, end date, end pos lat, end pos lon, id vehicle, kmh, start date, start pos lat, start pos lon, type vehicle. Figure 7 shows a very small time-frame (10 minutes) of car movements, this kind of graph may be used to find movement patterns and/or visually identify outliers.

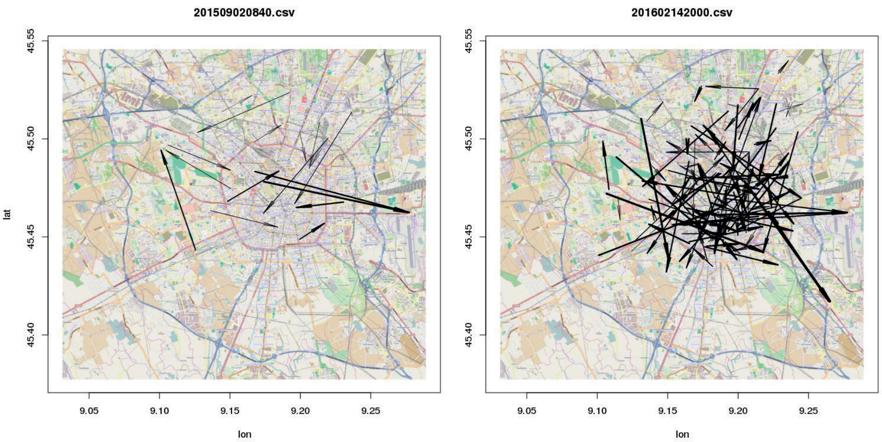


Fig. 7
At 2015/09/02 08:40 (left), two cars moving from the city centre towards the airport (rightmost arrows) in the morning;
at 2016/02/14 20:00 (right) many cars moving in the evening

3. Mandatory APIs proposal

The rationale of this proposal is based on these:
1) a resource (a large pool of cars) is converted from being private to "public" (private firms actually) control;
2) the private form of the resource is artificially limited (by regulations and policies) to push users towards the shared pool;
3) the cost burden is loaded on the public (because firms need to get revenues);
4) costs for users can be higher or lower than pre-sharing, based on usage profile;
5) workflow is heavily changed (from private to shared use);
6) environment impact can be lower if the whole system is well adapted to the actual needs of users;
7) at present, carsharing management is in the hands of private firms which do not publish enough data to let third parties implement independent analyses.

The authors' point of view is that some form of mandatory public access to status (both real-time and historical) data should be put into place by law, to give citizens the right to public accountability also in the field of carsharing. The ideal reification of this "third party" access would be the implementation of standardized APIs to be mandatorily supplied by carsharing vendors. This constraint should be introduced in every public tendering procedure.

Far from being a detailed and complete proposal (a standardization process - such as (Hovey and Bradner, 1996) - should be activated) a suggestion for such an API could be defined using a FSA (Finite State Automaton) representation of a single car in the carsharing system, using state transitions and dynamic and static properties: Dynamic: position (lat - long); fuel gauge; parked/taken anonymized user ID (if taken). Static: car ID; type (electric, diesel, gasoline, LPG, ...); odometer; usage fare; owner. A "start of discussion" function list proposal is listed below. The functions could be implemented via HTTP+REST

(Fielding and Taylor, 2000).

All the follwing functions should offer both the real-time and the historical version, i.e., they should also accept a "time-frame parameter such as in the last two examples:

- getStaticVehicles() => list of all "parked" vehicles;
- getMovingVehicles() => lista of all "non parked" vehicles;
- getInfo(id vehicle) => status of a single vehicle;
- getInfoService(service) => combined info on service
- getPath(id vehicle,interval) => returns the list of all paths followed by a vehicle in the time interval;
- getUserPath(id user,interval) => returns the list of all paths followed by a single user during the interval.

References

Roy T Fielding and Richard N Taylor. Architectural styles and the design of nework-based software architectures. University of California, Irvine Doctoral dissertation, 2000.

Jörg Firnkorn and Martin Müller. What will be the environmental effects of new free-floating carsharing systems? The case of car2go in Ulm. Ecological Economics, 70(8):1519–1528, 2011.

Richard Hovey and Scott Bradner. The organizations involved in the ietf standards process. 1996.

Diana Jorge and Gonçalo Correia. Carsharing systems demand estimation and defined operations: a literature review. European Journal of Transport and Infrastructure Research, 13(3):201–220, 2013.

Lester B Lave and Eugene P Seskin. Air pollution and human health, volume 6. Routledge, 2013.

John H Seinfeld and Spyros N Pandis. Atmospheric chemistry and physics: from air pollution to climate change. John Wiley & Sons, 2012.

Susanne Steinle, Stefan Reis, and Clive Eric Sabel. Quantifying human exposure to air pollution—moving from static monitoring to spatio-temporally resolved personal exposure assessment. Science of the Total Environment, 443:184–193, 2013.

Andrea Trentini. Lombardy epa obtorto collo data and anti-pollution policies fallacies. Journal of e-Learning and Knowledge Society, 10(2), 2014.

US-EPA. History of the clean air act. http://www.epa.gov/air/caa/amendments.html, 2013.

Tom Vanderbilt. Traffic: Why We Drive the Way We Do (and What It Says About Us). New York Times, 2009.

Authors

Andrea Trentini andrea.trentini@unimi.it

Computer Science PhD, Researcher at Dipartimento di Informatica (Computer Science Department) of Università degli Studi di Milano. He teaches "Programming 101", "Digital Citizenship and Technocivism" and "Embedded Systems". Free/Libre/Open Source Software advocate and knowledge sharing activist. He proudly defines himself a "hacker" (with the original term meaning). Founder of ArduinoAfternoon and HackLabCormano.

Federico Losacco federico.losacco@studenti.unimi.it

Computer scientist and student at Dipartimento di Informatica (Computer Science Department) of Università degli Studi di Milano. Very fond of tech-nology and Free/Libre/Open Source Software addict. He worked as a develo-per on embedded systems, mobile systems and web backend/frontend.