Why don't you drive my car?
Mining PSA car sharing data

Elisa Fromont, IRISA - Lacodam, Univ. de Rennes 1 elisa.fromont@irisa.fr
Alexandre Termier, IRISA - Lacodam, Univ. de Rennes 1 Alexandre.Termier@irisa.fr

Location: IRISA, team Lacodam, Rennes

Keywords: data science, data mining, trajectory pattern mining, machine learning, data visualization

Context and description of the internship:
Peugeot-Citroën (PSA) group's know-how encompasses all areas of the automotive industry, from production to distribution and services. Among others, its aim is to provide in many large cities, a car sharing service. This service consists in providing a fleet of cars and a "free floating" system which allows users to use a vehicle, then drop it off at their convenience in the city. To optimize their fleet and the availability of the cars throughout the city, PSA needs to analyze the trajectory of the cars and understand the mobility needs and behavior of their users. In the context of this internship, they will provide massive data coming from different cities (Madrid over 7 months and Berlin over 4 years) as well as some more data coming from a competitor, *Car2Go*, in Madrid and/or Washington. For their data, they can provide the position of 1000 different cars every 10s (which gives 60Mo/min of data).

Analyzing all this data requires skills in *Data Science.* Data science encompasses all the methods (data mining, machine learning, statistics, etc.) that help making sense of your data. The Lacodam team has a strong expertise in both data mining and machine learning [1]. The team's research favors works that advance the state of the art of this domain, while being useful for real world applications (such as retail industry [7], cybersecurity [8], human safety [9] or health [10]).

The aim of the internship is to survey the possible existing methods to tackle the real life problem proposed by PSA [2--6] and also investigate how some of the techniques developed in the Lacodam team (for example in emerging pattern mining or in deep learning) could be used to tackle the proposed problem. For example, it might be interesting to identify the behavioral patterns of users in cities where the car sharing system is well used compared to other cities where the service does not work as expected. It would also be interesting to predict the availability of cars in the fleet in the near future. The candidate will design a software (with the help of an INRIA engineer for the development part) to visualize and analyze the customer mobility behaviors and car trajectories.

If you like solving real life problems and learning much more about machine learning and data mining techniques, this internship is made for you.

Bibliography:
[1] U. Fayyad, G. Piatetsky-Shapiro and P. Smyth. **From Data Mining to Knowledge Discovery in Databases**. AI Magazine, 1996.

[2] H. Becker, F. Ciari, K. W. Axhausen. **Modeling free-floating car-sharing use in Switzerland: A spatial regression and conditional logit approach.** Transportation Research Part C: Emerging Technologies, Volume 81, August 2017, Pages 286-299.

[3] S. Weikl, K. Bogenberger. **Relocation Strategies and Algorithms for Free-Floating Car Sharing Systems.** IEEE Intelligent Transportation Systems Magazine, Volume: 5, Issue: 4, 2013.

[4] F. Ciari, B. Bock, M. Balmer. **Modeling Station-Based and Free-Floating Carsharing Demand-Test Case Study for Berlin.** Transportation Research Record: Journal of the Transportation Research Board V 2416, 2014, pages 37-47.

[5] J. Kopp, R. Gerike, K. W. Axhausen. **Do sharing people behave differently? An empirical evaluation of the distinctive mobility patterns of free-floating car-sharing members.** Transportation, 2015, Volume 42, Issue 3, pp 449–469.

[6] S. Herrmann, F. Schulte, and S. Voß. **Increasing Acceptance of Free-Floating Car Sharing Systems Using Smart Relocation Strategies: A Survey Based Study of car2go Hamburg**. International Conference on Computational Logistics ICCL 2014: Computational Logistics pp 151-162.

[7] C. Gautrais, R. Quiniou, P. Cellier, T. Guyet, A. Termier: **Purchase Signatures of Retail Customers**. Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD) 2017: pp 110-121

[8] A. Siffer, P.-A. Fouque, A. Termier, C. Largouët: **Anomaly Detection in Streams with Extreme Value Theory**. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD) 2017: pp 1067-1075

[9] K. Bascol, R. Emonet, E. Fromont and R. Debusschere: **Improving Chairlift Security with Deep Learning**,IDA'2017 Sixteenth International Symposium on Intelligent Data Analysis, to appear. Springer, London, 2017.

[10] Y. Dauxais, T. Guyet, D. Gross-Amblard, A. Happe: **Discriminant Chronicles Mining - Application to Care Pathways Analytics**. Conference on Artificial Intelligence in Medicine (AIME) 2017: pp 234-244