



10

Linking Complex Urban Systems: Insights from Cross-Domain Urban Data Analysis

Lelin Zhang, Bang Zhang, Ting Guo, Fang Chen,
Peter Runcie, Bronwyn Cameron, and Roger Rooney

Highlights While Open Data unlocks limitless opportunities for innovation, the potential could only be fully realised by a comprehensive understanding of data from multiple urban sub-systems.

L. Zhang (✉) • T. Guo • F. Chen
Data61 CSIRO, Eveleigh, NSW, Australia

University of Technology Sydney, Ultimo, NSW, Australia
e-mail: lelin.zhang@uts.edu.au; ting.guo@data61.csiro.au; fang.chen@uts.edu.au

B. Zhang • P. Runcie
Data61 CSIRO, Eveleigh, NSW, Australia
e-mail: matt.zhang@data61.csiro.au; peter.runcie@data61.csiro.au

B. Cameron
Sydney Water, Parramatta, NSW, Australia
e-mail: bronwyn.cameron@sydneywater.com.au

R. Rooney
ACT Government, Canberra, ACT, Australia
e-mail: roger.rooney@act.gov.au

- Various machine learning and data analytics techniques have been used to address the unprecedented complexity of cross-domain urban data.
- Individual industries could expand the horizon of their data analytics projects by adopting inter-domain data.
- Policymakers can make more informed and comprehensive decisions with a holistic view of urban systems.
- There are still significant challenges in making organisation-owned data open to public. But potential solutions exist and the related techniques are fast developing.

Introduction

Urbanisation is a global trend that has resulted in a fast-growing demand for infrastructure, transportation, energy, dwelling, education, healthcare, entertainment, communication, banking and financing and other urban services. Cities are complex socioeconomic systems that must provide all these services. Urban systems or cities consist of a large number of sub-systems which must support individual demands and also interact closely with each other. A comprehensive understanding of the whole urban system can help us make efficient urban asset maintenance, accurate forecasts and informed plans for future demand.

However, in the past, data about different urban sub-systems are often locked up in different government agencies and organisations. Studies of a sub-system usually can only access isolated and fragmented datasets, and therefore treat the sub-system in an isolated and fragmented way, rather than considering it holistically. Fortunately, following the recent Open Data advocacy, more and more public data about urban sub-systems have been released by government agencies and organisations. This unlocks limitless opportunities for all parties in the urban systems to collaborate and innovate, not only for individual sub-systems but also for the urban system as a whole.

Along with the Open Data movement, studies are now aiming to “move from data to information to knowledge, and, ultimately, to action for urban sustainability and human well-being.” (Ramaswami, Russell, Culligan, & Karnamadakala Rahul Sharma, 2016). To derive insights from the data collected in disparate areas, we need cross-domain data-driven solutions. In this chapter, we will shed light on three successful cases on water utility, parking and urban planning projects. This showcases how cross-domain urban data analysis techniques, combining a mixture of public urban data and private sub-system-specific data, can reshape the decision-making processes of businesses, governments and societies.

While current projects are often driven by a public/private organisation of a closed nature, we believe the general public and the organisations themselves could benefit from a more open environment. Following the case studies, we conclude this chapter with discussions on possible directions to increase the openness of urban sub-system data analytic projects, and how such movements could be a win-win for all parties.

Urban Wastewater Pipe Blockage Prediction

Background

For water utilities, wastewater pipe (sewer) blockages pose a great challenge to the daily operation of wastewater pipe networks. Not only do they result in high economic costs in emergency repairs and clean up, but they also have social impacts, such as service disruption, environmental pollution and road and amenity closure. Traditional forecasting and prediction techniques are currently not capable of accurately predicting sewer chokes. Sydney Water, one of the largest water utilities in Australia, sponsored a study with us to address this challenge. The aim was to investigate and understand the key factors that impact the sewer blockage patterns and develop a prediction model to predict future blockage probability. The outcome of this project provides data-driven decision support to water utilities that could lead to a more efficient and predictive maintenance strategy.

If current field validation of the model is successful, a preventative maintenance programme may reduce the social, reputational and environmental impact of sewer chokes, enable more efficient allocation of resources, improve regulatory compliance and reduce disruptions, rebates and property damage costs.

Interestingly, the data from vegetation coverage, climate, soil and demography plays a significant role in the solution, which demonstrates the value of cross-domain data analysis.

Datasets

Sewer blockages can be caused by many things, including tree roots, grease, debris and foreign objects (e.g., soft wipes). There are many factors which may contribute to blockage incidents, including the pipe's intrinsic characteristics (diameter, length, material, etc.), the pipe's external environment (vegetation coverage, climate and soil condition, property type and demographical characteristics) and historical blockage events. In order to achieve accurate prediction, we investigated the following datasets (Table 10.1).

Table 10.1 Datasets used for urban wastewater pipe blockage prediction

Dataset	Description	Source
Sewer network	Characteristics of pipes, including laid year, diameter, length, material, location, etc.	Private—water utility
Blockage records	Historical blockage records, including blocked pipe, date, type of blockage, blockage location, etc.	Private—water utility
Vegetation coverage	Shapes (polygons) of tree canopies for more than 4 million trees, obtained from satellite imagery.	Private/Public—third party
Climate and soil	Rainfall, temperature, evaporation and soil moisture.	Private/Public—Bureau of Meteorology and Office of Environment and Heritage ^a
Demographic	Property types and densities, population, etc.	Public—Australian Bureau of Statistics ^b

^a<http://www.bom.gov.au/climate/data/>

^b<http://www.abs.gov.au/websitedbs/censushome.nsf/home/datapacks>

Factor Analysis

Based on the historical data of sewers and blockages, we conducted statistical analysis (Li, Zhang, Wang, Chen, & Vitanage, 2014; Li et al., 2015) to discover the influences of different factors. This provides useful insights towards the prediction model, such as:

Pipe characteristics: using only the sewer and blockage data, we could get the insights about the performance of pipes with different characteristics. Figure 10.1 shows the blockage rates of pipes laid in different years. Figure 10.2 shows the blockage rates for a few common materials, which suggests that pipes made of PVC (polyvinyl chloride) are less prone to blockage than other materials. When local knowledge of the number and types of pipes laid in different years is taken into account, these confirm the domain knowledge held by experts in water utility.

Tree coverage: tree root intrusion is one of the most common causes of sewer blockages in the Sydney Water network. For a tree, the extent of its root system is dependent on the species, age, nutrient availability and physical limitations of surrounding soil. However, the tree canopy often provides a good indication of the size of the root system. Sydney Water, in collaboration with Jacobs, developed a tree canopy polygon layer in order to understand the relationships between vegetation and blockages caused

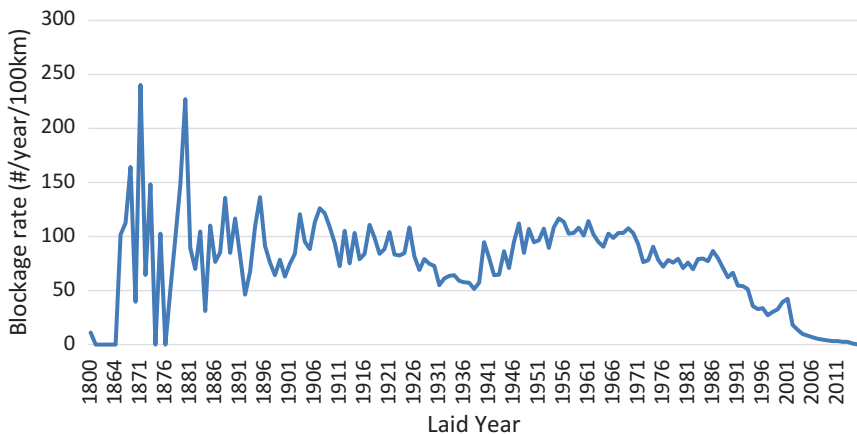


Fig. 10.1 Blockage rates of wastewater pipes laid in different years

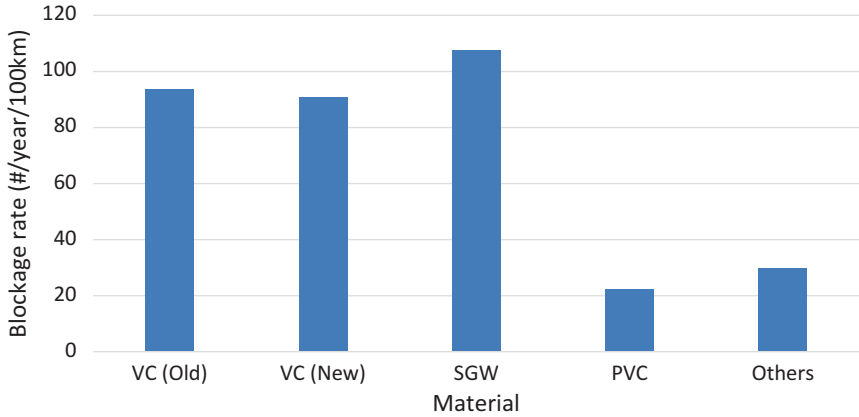


Fig. 10.2 Blockage rates for wastewater pipes made of different materials



Fig. 10.3 Tree canopy polygons

by tree roots. The pipe network is overlaid on the tree canopy polygons (refer to Fig. 10.3) to compute the tree canopy coverage percentage of each pipe. As shown in Fig. 10.4, a clear positive correlation between tree canopy coverage and blockages caused by tree roots can be observed.

Climate and soil conditions: to understand the impact of climate and soil condition to sewer blockages, we matched the temperature, rainfall, evaporation and soil moisture data to each pipe in the whole network. A strong correlation between the climate and soil conditions

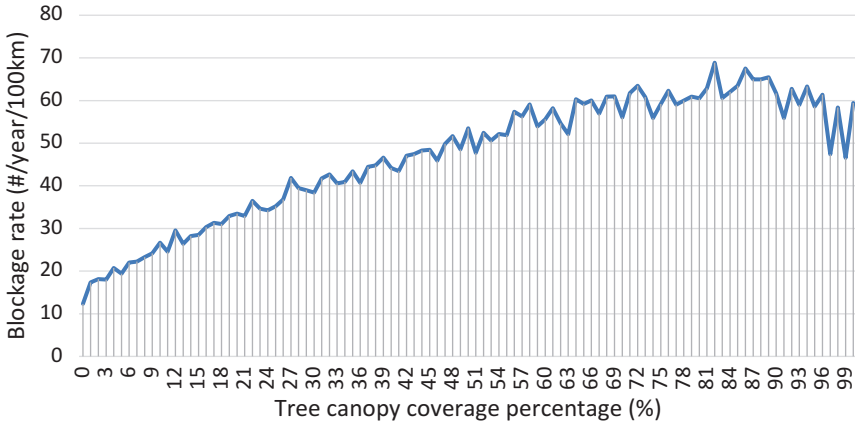


Fig. 10.4 Tree root caused blockage rates under different tree canopy coverage percentages

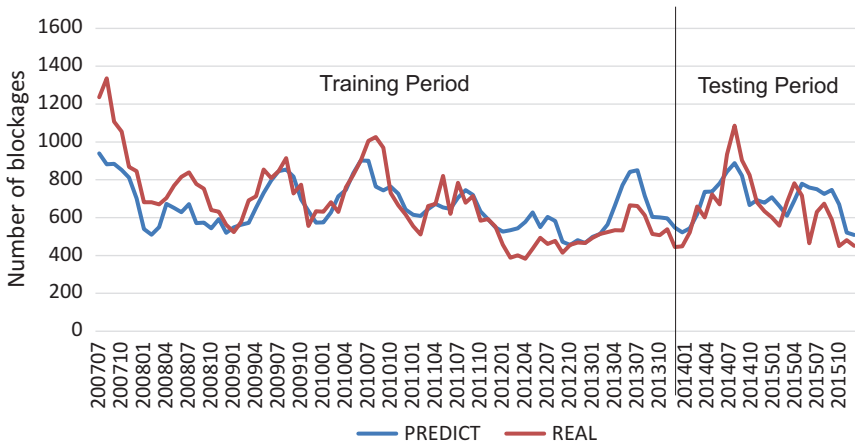


Fig. 10.5 Predictive power of climate factors (temperature, rainfall, evaporation and soil moisture) for predicting tree root caused blockages six months later

and the blockages occurred 6 months later is observed. One theory is that the climate in summer affects the growth of tree roots, which in turn affects the number of blockages for the following winter peak. Figure 10.5 demonstrates the predictive power of the climate and soil conditions on the blockages.

Methodology and Outcome

We proposed a new variant of Hawkes process (Hawkes, 1971) for short-term (one year) sewer blockage prediction (Lin et al., 2015, 2016a). Hawkes process is a stochastic point process (Daley & Vere-Jones, 2002) based statistical model. In this study, each blockage event is treated as a point, and the point intensities (number of blockages in the future) are predicted using both background intensity and trigger intensity. The background intensity models the impact of a pipe’s intrinsic characteristics and external environment on blockage behaviours, while the trigger intensity models the contribution of a pipe’s past blockage history on future blockage events. All the key influential factors discovered in the factor analysis phase are utilised in the prediction model.

To evaluate the performance of our predictive model, we split the 15-year blockage data into two parts: the first 14 years’ data are used to train the model, and the model is blind-tested using the last year’s data. As shown in Fig. 10.6, the prediction model could accurately predict risk of blockages in testing year: the top 1% pipes with high-predicted blockage risk contribute to about 8% of the blockages in the testing year. In other words, if top 1% high-risk pipes according to prediction were fixed, about 8% of the blockages would have been prevented. Similarly, fixing 10% of the pipes according to

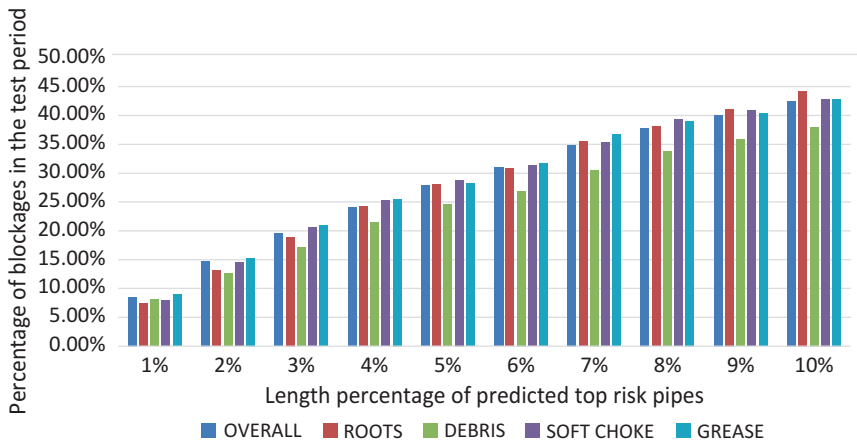


Fig. 10.6 Blockage prediction curve, showing the contributions of blockages in the test period for predicted top risk pipes

the predicted risk would have prevented more than 40% of blockages. Sydney Water is conducting a field validation to confirm the performance of the model, and if passed, the predictive model could improve the efficiency and reduce the cost of the future maintenance schedules.

This project illustrates how water utilities could improve their maintenance schedules by bringing in open urban data. With the predictive model driven by cross-domain data, water utilities could reduce their operation cost and service interruptions, providing a better service with lower cost to the public. Modern cities are a complex system with sub-systems interacting with each other. With the movement of Open Data, datasets of different sub-systems will be empowering similar techniques in other asset-intensive urban sub-systems, bringing better services to the citizens.

Smart Parking Occupancy Pattern Analysis

Background

With the increasing demand and expectation for public parking spaces, smart parking systems are being adopted in many parking lots. They utilise sensing devices to monitor occupancies in real-time, and distribute the information via various channels, such as signboards, Internet websites and mobile applications. They help reduce parking time, ease traffic load burden and better utilise the parking spaces.

In this study, we collaborated with a local government agent on a trial of smart parking system. The aim is to understand the parking patterns from the sensor data, and develop a prediction model for the future occupancy rates. The outcomes of the data analysis and prediction model provide decision support for both policymakers and motorists.

Datasets

The parking lots in this trial are located at a local central business area, surrounded by shops, restaurants, a cinema and a cricket/oval ground hosting Australian Football League (AFL) games. Each of the facilities

Table 10.2 Datasets used for smart parking occupancy pattern analysis

Dataset	Description	Source
Parking records	Records of more than 4 million parking events for more than 400 bays, including start time, end time, bay, lot type (1P Free, 2P Voucher, etc.), location, etc.	Public—Open Data Portal (ACT Government, 2017) ^a
Public Wi-Fi	Number of users connected to nearby public Wi-Fi access points.	Private—Local government
Payment transactions	Payment records for paid parking lots.	Private—Local government
Special days	Dates of public holidays and events (Father's day, Mother's day, AFL game).	Public—Public
Weather	Temperature, rainfall and solar.	Public—Bureau of Meteorology ^b

^a<https://www.data.act.gov.au/Transport/Smart-Parking-Stays/3vsj-zpk7>

^b<http://www.bom.gov.au/climate/data/>

has different attendance patterns, which generates different parking demand. In order to understand the impacting factors for the parking patterns, we investigated the following datasets (Table 10.2).

Factor Analysis

In order to identify influential factors for predicting parking patterns, we conduct statistical analysis on different factors. We analysed the parking patterns in three metrics:

- Occupancy rate: the utility ratio of a parking bay, defined as the percentage of duration that the bay is occupied by vehicles.
- Arrival rate: the turn-over of a parking bay, defined as the number vehicles arrived per hour per bay.
- Overstay rate: the percentage of vehicles that stayed more than 10 minutes over the parking time limit.

A few useful insights discovered include:

Hour of day: as one would expect, the demand for parking spaces would have its peak and non-peak time during a day. Table 10.3 summarises the

Table 10.3 Occupancy rate for different hour of day and different lot types

Hour	Overall	5Min Free	1/4P Free	1/2P Free	1/2P Voucher	1P Free	1P Voucher	2P Free	2P Voucher	4P	8P	Disabled	Dropoff	EV	LZ
0	0.053	0.119	0.043	0.076	0.079	0.094	0.114	0.126	0.042	0.022	0.030	0.016	0.067	0.013	0.025
1	0.043	0.095	0.036	0.052	0.050	0.081	0.081	0.122	0.032	0.019	0.028	0.012	0.038	0.008	0.016
2	0.042	0.073	0.039	0.039	0.038	0.078	0.082	0.148	0.029	0.018	0.028	0.010	0.022	0.007	0.016
3	0.043	0.064	0.034	0.032	0.034	0.074	0.104	0.166	0.027	0.018	0.028	0.010	0.021	0.007	0.014
4	0.045	0.060	0.037	0.029	0.035	0.069	0.120	0.175	0.027	0.019	0.028	0.010	0.015	0.010	0.014
5	0.062	0.069	0.046	0.029	0.063	0.076	0.199	0.205	0.058	0.019	0.027	0.018	0.016	0.016	0.015
6	0.107	0.224	0.067	0.033	0.148	0.157	0.270	0.232	0.158	0.021	0.028	0.065	0.023	0.016	0.029
7	0.200	0.634	0.339	0.070	0.284	0.337	0.367	0.308	0.357	0.028	0.039	0.120	0.036	0.017	0.067
8	0.308	0.720	0.457	0.452	0.317	0.521	0.402	0.451	0.573	0.050	0.105	0.219	0.043	0.043	0.137
9	0.430	0.776	0.552	0.776	0.419	0.620	0.489	0.653	0.760	0.121	0.202	0.398	0.090	0.097	0.272
10	0.567	0.790	0.625	0.858	0.606	0.769	0.612	0.774	0.897	0.308	0.246	0.691	0.116	0.296	0.422
11	0.634	0.827	0.680	0.885	0.714	0.833	0.671	0.773	0.912	0.450	0.293	0.811	0.162	0.409	0.468
12	0.725	0.863	0.734	0.926	0.874	0.934	0.742	0.784	0.946	0.646	0.334	0.869	0.327	0.521	0.492
13	0.729	0.853	0.735	0.910	0.846	0.897	0.733	0.770	0.929	0.706	0.357	0.788	0.339	0.516	0.446
14	0.646	0.837	0.728	0.865	0.739	0.805	0.676	0.723	0.867	0.536	0.339	0.717	0.236	0.443	0.360
15	0.579	0.841	0.739	0.867	0.689	0.735	0.675	0.679	0.831	0.393	0.289	0.625	0.243	0.371	0.286
16	0.539	0.844	0.760	0.854	0.725	0.731	0.695	0.640	0.770	0.335	0.237	0.508	0.237	0.345	0.289
17	0.543	0.859	0.764	0.831	0.873	0.804	0.762	0.540	0.739	0.383	0.148	0.444	0.376	0.419	0.478
18	0.653	0.870	0.735	0.883	0.917	0.913	0.811	0.530	0.888	0.643	0.108	0.550	0.501	0.596	0.691
19	0.694	0.870	0.678	0.885	0.913	0.920	0.792	0.523	0.902	0.761	0.160	0.592	0.540	0.625	0.723
20	0.605	0.847	0.574	0.792	0.868	0.830	0.739	0.451	0.759	0.646	0.147	0.455	0.483	0.550	0.647
21	0.404	0.764	0.414	0.538	0.729	0.607	0.601	0.316	0.466	0.371	0.094	0.243	0.352	0.375	0.510
22	0.232	0.594	0.212	0.306	0.476	0.399	0.431	0.231	0.221	0.181	0.059	0.121	0.218	0.197	0.221
23	0.113	0.224	0.067	0.146	0.219	0.203	0.262	0.179	0.090	0.072	0.039	0.045	0.115	0.053	0.069

occupancy rate for different lot types. Two peak times for lunch (11:00–13:00) and dinner (18:00–20:00) are observed for most lot types.

Special days: we matched the parking records with the dates of public holidays and events to see if there are any differences in parking patterns. As depicted in Fig. 10.7, such days do exhibit different parking patterns compared to a normal day, and each day has its unique pattern. For instance, the occupancy rate is higher on Easter Saturday, AFL game day and Mother’s Day, but much lower on Christmas and Boxing Day.

Weather: to investigate the parking behaviours in different weather, we matched the parking records with the local temperature, rainfall and solar records. However, no significant difference in parking pattern is observed. Figure 10.8 shows the results for rainfalls, no matter a day is sunny or rainy, the occupancy, arrival and overstay rate stay in same levels.

Methodology and Outcome

In order to help motorists plan their trips, we build an ensemble-based regression model (Polikar, 2006) that takes all the factors as input to predict next hour occupancy rate for every parking lot. The model is chosen

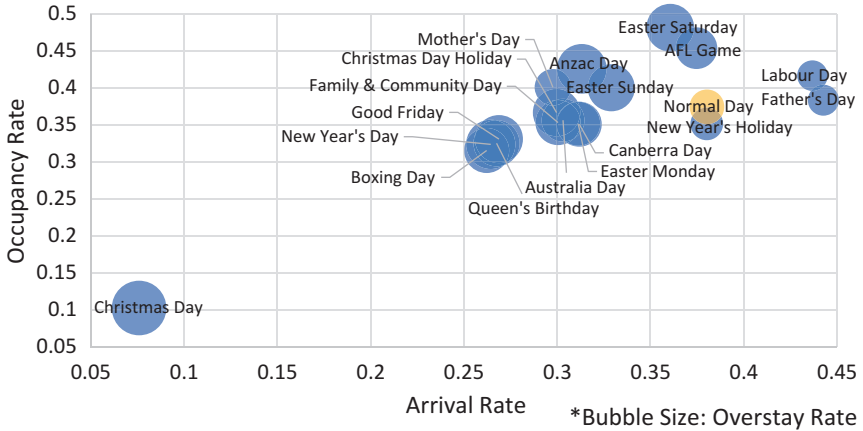


Fig. 10.7 The occupancy rate, arrival rate and overstay rate for different public holidays/events

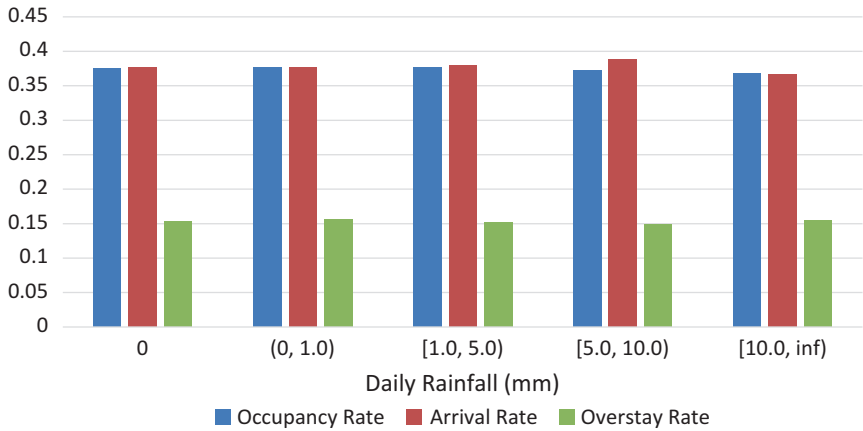


Fig. 10.8 The occupancy rate, arrival rate and overstay rate on days with different rainfall

because it is able to handle heterogeneous data types from different data sources, and it offers competitive prediction performance.

The predictive model is tested by using the standard ten-fold cross-validation process. That is, the data are split into ten folds, and the model is trained and tested for ten rounds, where in each round, nine folds of data are used to train the model, and one fold is left to test the results.

The overall R^2 score is 89.9%, which means the model could predict the occupancy rate with high accuracy, explaining about 90% of the variation. According to a survey for the trial smart parking app, most users find the occupancy rate prediction function to be useful in planning their trips and reducing parking time.

The project illustrates how Open Data could be contributing to the modelling of parking availability. Like other urban sub-systems, the traffic condition of a city is impacted by many factors. By providing a better guide, the parking time is reduced, and the traffic condition improved. To bring greater good to the community, and to inspire further research, the parking data is made available on the Open Data Portal (ACT Government, 2017).

Urban Function and Region Popularity Analysis

Background

In recent years, the fast-growing urban population and shifting urban functions pose a major challenge for urban planners. In a joint study with the planning department of a regional government, we aim to understand various factors involved in these processes, and answer the question: “where, when and why a region develops?” (Zhang, Zhang, Guo, Wang, & Chen, 2018)

Datasets

There are many factors influencing the developments of urban systems. In this study, we investigated the following datasets (Table 10.4).

Methodology and Outcome

We performed three different analyses on the collected cross-domain datasets:

Dwelling productions: dwelling production is an important indicator of urban development. Using the macroeconomic indicators as input factors,

Table 10.4 Datasets used for urban function and region popularity analysis

Dataset	Description	Source
Development application	Time, location, type, scale of new developments.	Public—Planning department ^a
Macroeconomic indicators	GDP, inflation rate, labour markets, housing, population, market/consumer performance, etc.	Public—Australian Bureau of Statistics and Reserve Bank of Australia ^b
Urban functions	More than 100 categories of business and public facilities, including location, type and size.	Public—Google Places API ^c
Properties	Type (house/apartment/unit), configurations (number of bedrooms), historical price.	Public—real estate websites ^d

^a<https://data.gov.au/dataset/sydney-region-dwellings/resource/cd918b27-05a8-426e-9d9e-e54d5b963c53>

^b<https://www.rba.gov.au/statistics/tables/>

^c<https://developers.google.com/places/web-service/intro>

^d<https://www.realestate.com.au/>

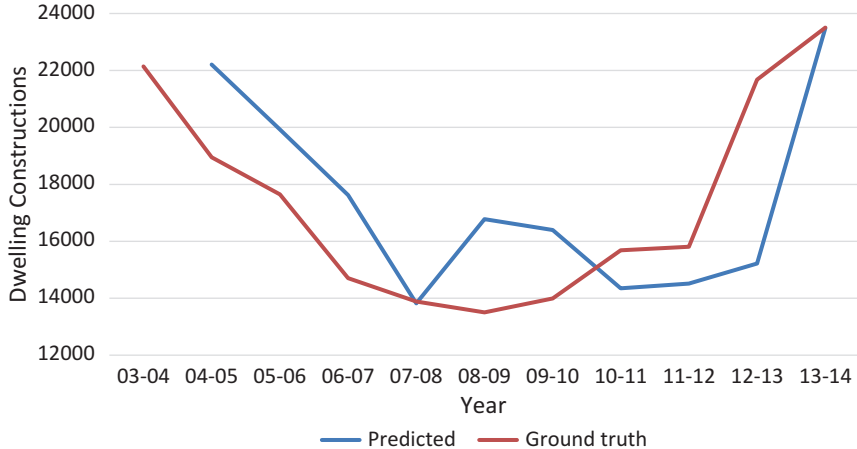


Fig. 10.9 Predictive power of macroeconomic indicators on dwelling constructions (DC)

we aim to predict when and where new dwelling will appear. As shown in Fig. 10.9, macroeconomic indicators are informative predictors for the dwelling production, yielding an accurate prediction.

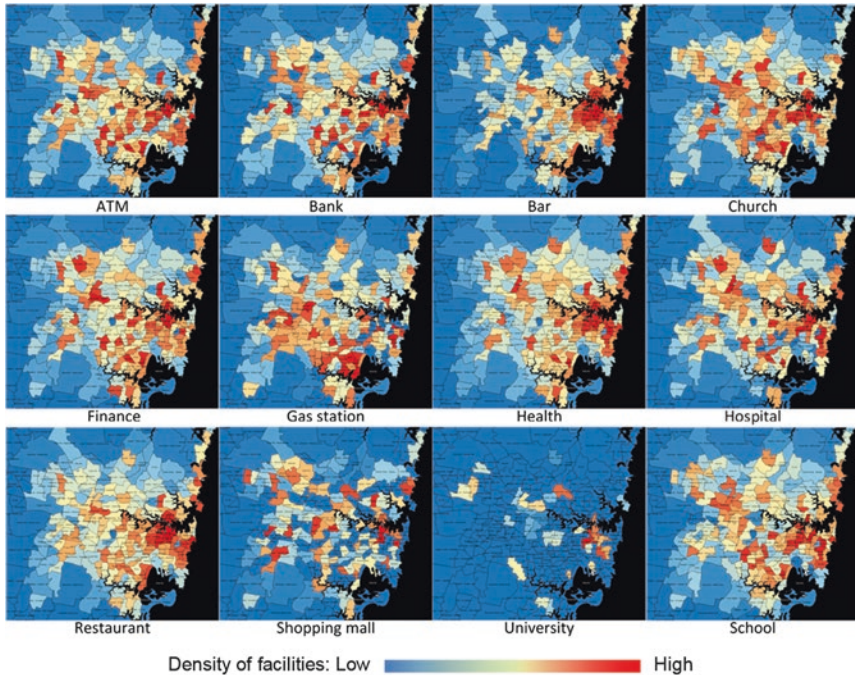


Fig. 10.10 Examples of urban functions for the greater Sydney area

Urban functions: we try to understand the similarities and differences of different urban regions in terms of urban functions. To achieve this, we first compute the function intensities for each region from the business and public facility location data. The function intensity for a particular type of business and public function (e.g., bank, restaurant or school) measures the density of facilities with that function. Some examples of urban functions are shown in Fig. 10.10.

With the function intensities for each region, we then cluster the regions into different groups using a Bayesian nonparametric mixture model with spatial constraints (Lin et al., 2016b). The algorithm offers a tool to discover the regional patterns of urban functions and the similarities among regions.

Region popularity: lastly, we combine the urban functions and property data to analyse the region popularity. A self and mutually excited stochastic interaction point process (Lin et al., 2016a, 2016b) is used to

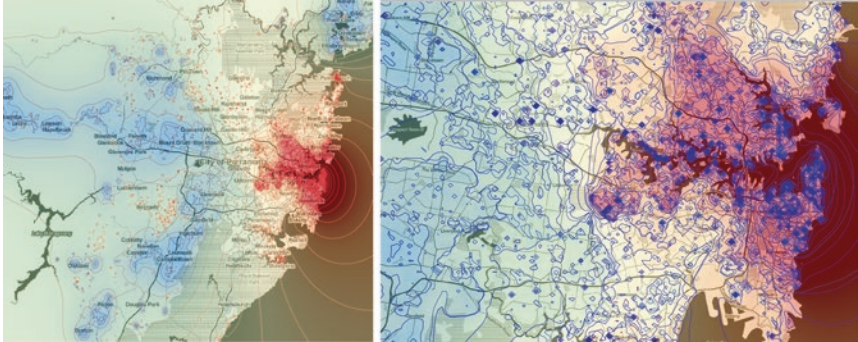


Fig. 10.11 Region popularity (median property price) heat map and evaluation contour line for the greater Sydney area

estimate the region popularity based on both the historical property price and region functions. The estimated popularity heat map and price contour lines are shown in Fig. 10.11. The model (Zhang et al., 2018) provides a useful tool to understand the interactions between urban functions and regional popularities.

Conclusion and Future Direction

The advocacy of Open Data helps make a huge amount of valuable urban data publicly available. The potential of such data, however, can only be fully realised by adopting a holistic cross-domain view of them. In many cases, the open urban data needs to be exploited with the support of organisation-owned private data for maximising its value. In this chapter, we showcased how industries and government agencies can utilise a combination of cross-domain data from both public and private sources to improve their capabilities in planning, operating and maintenance. These successful projects demonstrate the demand, methodology and benefit of the cross-domain urban data analysis. In the following, we share some of our observations and considerations on how to fully release the power of open urban data and how to gradually make more privately owned urban data publicly open for the society to use.

- Thanks to the advanced data analytic techniques, especially the recent advance in artificial intelligence and machine learning, the value of open urban data can be greatly released and help generate valuable insights for supporting efficient urbanisation.
- Synergy exists between open and private urban data. Open urban data can help leverage the potential value of organisation-owned private data. Both open and private urban data can work together to empower the efficient operation and maintenance of urban systems.
- Urban system has many sub-systems, for example, urban infrastructure, urban transportation and so on, which are working closely with each other. Making one sub-system's data public open can benefit other sub-systems via cross-domain urban data analysis.
- Due to the data security and privacy concerns, it is difficult for many industrial organisations to make their datasets public open. However, the value of such private urban data is significant. Hence, efforts need to be made to help industrial organisations to gradually release their data and make it publicly available for other parties' utilisation.
- Research community has also paid attention to help make private data safely open for sharing. For instance, privacy preserving data analysis and machine learning on encrypted data techniques (Aldeen, Salleh, & Razzaque, 2015; Graepel, Lauter, & Naehrig, 2013) have attracted increasing attention and have been widely adopted in many different areas for making private data safely sharable.

Performing cross-domain data analysis for a single urban sub-system is just the first step. Looking into the future, more efficient optimisation of urban systems could be achieved if we could jointly optimise multiple sub-systems together. For example, by considering the current demographics and new developments of shops and dwellings, local governments could obtain an accurate forecast of how the urban functions will change in the future, and make collaborated policies to facilitate such transition. Such information could also be passed to utility and telecommunication companies to ensure the infrastructures could be built to match the growing demand. We believe that Open Data, combined with cross-domain data analysis, could transform future cities into smart, inclusive and responsive open cities.

Acknowledgements The authors would like to thank our Sydney Water collaborators, Bronwyn Cameron, Mark McGowan, Craig Mitchell, Judith Winder and Rod Kerr, for the wastewater pipe blockage prediction work.

References

- ACT Government. (2017). *Smart parking stays | Open data portal*. [Online]. Retrieved May 21, 2018, from <https://www.data.act.gov.au/Transport/Smart-Parking-Stays/3vsj-zpk7>
- Aldeen, Y. A. A. S., Salleh, M., & Razzaque, M. A.. (2015). *A comprehensive review on privacy preserving data mining*. s.l.: SpringerPlus.
- Daley, D., & Vere-Jones, D. (2002). *An introduction to the theory of point processes*. s.l.: Springer.
- Graepel, T., Lauter, K., & Naehrig, M. (2013). *ML confidential: Machine learning on encrypted data*. s.l.: Information Security and Cryptology—ICISC 2012. ICISC 2012. Lecture Notes in Computer Science, vol. 7839. Springer, Berlin, Heidelberg.
- Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1), 83–90.
- Li, B., Zhang, B., Li, Z., Wang, Y., Chen, F., & Vitanage, D. (2015). *Prioritising water pipes for condition assessment with data analytics*. s.l.: OzWater.
- Li, Z., Zhang, B., Wang, Y., Chen, F., & Vitanage, D. (2014). *Water pipe condition assessment: A hierarchical Beta process approach for sparse incident data*. s.l.: Machine Learning.
- Lin, P., Zhang, B., Wang, Y., Li, Z., Li, B., Wang, Y., & Chen, F. (2015). *Data driven water pipe failure prediction: A Bayesian nonparametric approach* (pp. 193–202). New York, NY: ACM International Conference on Information and Knowledge Management.
- Lin, P., Zhang, B., Guo, T., Wang, Y., & Chen, F. (2016a). *Infinite hidden semi-Markov modulated interaction point process* (pp. 3900–3908). Barcelona, Spain: Advances in Neural Information Processing Systems.
- Lin, P., Zhang, B., Guo, T., Wang, Y., & Chen, F. (2016b). *Interaction point processes via infinite branching model*. Phoenix, AZ: Thirtieth AAAI Conference on Artificial Intelligence.
- Polikar, R. (2006). Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, 6(3), 21–45.

- Ramaswami, A., Russell, A. G., Culligan, P. J., & Karnamadakala Rahul Sharma, E. K. (2016). Meta-principles for developing smart, sustainable, and healthy cities. *Science*, 940–943.
- Zhang, B., Zhang, L., Guo, T., Wang, Y., & Chen, F. (2018). *Simultaneous urban region function discovery and popularity estimation via an infinite urbanization process model*. London, UK: ACM SIGKDD Conference on Knowledge Discovery and Data Mining.