An Intelligent Predictive Analytics System for Transportation Analytics on Open Data Towards the Development of a Smart City

Abdul-Rasheed A. Audu[1], Alfredo Cuzzocrea[2], Carson K. Leung[1(✉)], Keaton A. MacLeod[1], Nibrasul I. Ohin[1], and Nadège C. Pulgar-Vidal[1]

[1] University of Manitoba, Winnipeg, MB, Canada
kleung@cs.umanitoba.ca
[2] University of Trieste, Trieste, TS, Italy

Abstract. As time is a precious asset, bus riders would desire to get accurate information about bus arrival time. Although different research approaches have been developed to correctly predict bus arrival time, very few of them produce highly precise and accurate results based on open data. In this paper, we present an intelligent system designed for transportation analytics on open data such as bus delay data. Specifically, the system accesses open data to analyze public transport data—such as historical bus arrival time—for urban analytics; it then conducts data analytics and mining to discover frequent patterns. Based on the discovered patterns, the system makes predictions on whether the bus arrives on time or is being late. Evaluation on real-life open data provided by a Canadian city show the effectiveness and prediction accuracy of our intelligent system in transportation analytics on open data. The results are encouraging towards the goal of developing smart cities.

Keywords: Intelligent system · Transportation analytics · Open data · Public transportation · Bus · Bus delay · Data analytics · Frequent pattern mining · Predictive analytics · Smart city

1 Introduction

As we are living in the era of big data [1–5], big data are everywhere. With advances in technologies, huge volumes of a wide variety of valuable data—which may be of different levels of veracity (e.g., precise data, imprecise and uncertain data)—can be easily generated or collected at a high velocity. They can be originated from a wide variety of data sources in various real-life applications. These include Internet of Things (IoT) data [6], music [7], stock prices [8], meteorological data [9], web data [10], and urban data (e.g., public transit data).

Public transit has become the daily mode of commuting for a large portion of the population due to its convenience in many scenarios [11, 12]. This includes the low cost of commuting (cf. owning and operating a private vehicle), as well as the environmental effect perceived by individuals who are concerned about the amount of carbon being emitted into the atmosphere daily [12, 13]. Due to the ease of data storage techniques, data relating to transit usage and transit delays have been collected and

L. Barolli et al. (Eds.): CISIS 2019, AISC 993, pp. 224–236, 2020.
https://doi.org/10.1007/978-3-030-22354-0_21

made accessible in the form of open data in multiple cities across the world. For instance, in Canada, from (Pacific) coast to (Atlantic) coast, many Canadian provinces and territories, as well as municipalities, have joined the conversation on *open government*, in which residents can find and easily access information about open government activities in their jurisdiction and/or across the country. Specifically, so far, nine provincial governments and 66 municipal governments have set up open data portals. Examples include the following:

- City of Toronto Open Data Portal[1], and
- City of Winnipeg Open Data Portal[2].

Among the available data in these open data portals, we focus on transit data in this paper as transit data are closely related to most residents. In particular, we examine bus delay data[3] from Toronto Transit Commission (TTC).

In general, *transit delay* is a regular occurrence and its patterns take an interesting form [14]. In this paper, we analyze bus delay patterns over time, study the patterns that exist, and observe delay correlations which can be used to predict the possibility of a future delay given previous scenarios. These results can give an insight into existing problems in the infrastructure, bus routes and bus schedules; they can also give an insight into future planning and development of a safer, more rider-friendly, more rider-convenient transit systems.

The transit system is an integral part to the lives of many people in modern society. It facilitates a safe way for people to move across their cities for a variety of reasons (e.g., commuting to work, school, or shopping mall; visiting friends; going to recreation centers, sports games, or concerts). It also provides services for both frequent bus riders (who usually hold weekday passes, weekly passes, or monthly passes) and occasional bus riders.

As seasons come and go, weather changes, and events of different scales happening in the cities, there may be problems of maintaining the schedule of some buses due of one factor or a combination of factors [14]. To prevent and reduce the number of delays that occur, knowing (a) the causes of delays and (b) the delay trends that appear frequently can serve as a starting point for policymakers in the affected cities. Transit riders are limited to knowing only when and where their buses are supposed to arrive, without any other reliable information. In this paper, our *key contribution* is our intelligent system for analyzing public transport data on bus arrival time and predicting on whether or not bus arrives on time or late. Specifically, our system conducts frequent pattern mining for predictive analytics on open data for transportation analysis. To evaluate our intelligent system, we apply it to real-life bus delay data provided by the TTC. It predicts future bus delay data based on the frequently mined patterns from previous years. Evaluation results show that the system accurately predicts bus delay data for test data (for year 2017) based on the frequent pattern mining of the historical bus delay data (from years 2014–2016).

1 https://www.toronto.ca/city-government/data-research-maps/open-data/open-data-portal/.

2 https://data.winnipeg.ca/.

3 https://portal0.cf.opendata.inter.sandbox-toronto.ca/dataset/ttc-bus-delay-data/.

The remainder of this paper is organized as follows. Next section discusses related works. Section 3 describes our intelligent system for transportation analytics of bus data and predictive analytics for bus delay. Evaluation on Sect. 4 shows the efficiency of our intelligent system in predicting bus delay for real-life situations for buses operated by TTC. Finally, Sect. 5 draws conclusions.

2 Related Works

There has been works on predictive analytics [15–19] on various real-life applications. In particular, there has been research works done to accurately predict bus arrival times. For instance, Sun et al. [20] focused on predicting the correct bus arrival time based on the geographic information system (GIS)-based map-matching algorithm, which is used to project each received location onto the underlying transit network, thus pinpointing bus location to predict its arrival. Lin et al. [21] also focused on predicting the correct bus arrival times, but based on global positioning system (GPS) data and automatic fare collection (AFC) system data with the help of *artificial neural network* (*ANN*). Rajput et al. [12] analyzed New York open data, and applied *clustering* to identify (a) highly congested areas and (b) areas with less bus stops to provide suggestion for bus stops. Many other researchers [22–26] used *Kalman filter* or *time series models* to predict future travel times, under the assumption of a direct relationship with previous travel times. In other words, most of the aforementioned approaches

- use auxiliary information such as GIS, GPS, and/or AFC system data (which may not be easily accessible by general public); and/or
- apply data mining tasks like classification and/or clustering—via techniques like ANN, Kalman filter, and/or time series models.

In contrast, in our current paper, we use *open data* (which are freely accessible by general public) and apply a different data mining task of *frequent pattern mining* (e.g., using the FP-growth algorithm) on the bus delay data of Toronto to generate a predictive model to predict possible bus delays based on day and time of the delay, delay duration, delay severity, and delay type.

3 Construction of Our Intelligent System

To conduct transportation analytics on open data (specifically, predictive analytics on bus delay data), our intelligent system first cleans the input data. It then analyzes the cleaned data to discover interesting patterns and to make accurate predictions.

3.1 Data Understanding

The TTC provides delay data for the following modes of ground transportation served in the City of Toronto:

- buses,
- streetcars (i.e., trams, trolleys), and
- subways (i.e., underground rapid transit rails).

In the current paper, we focus on the bus data. However, it is important to note that the knowledge learned from the current paper on TTC bus delay data can be transferred to delay data for other transportation modes such as streetcar and/or subway—via *transfer learning*. Along the same direction, such knowledge can also be transferred to other jurisdictions and/or other transportation data via transfer learning.

For bus delay data (such as the TTC bus delay data available on the City of Toronto Open Data Portal), they usually contain the following information:

- report date, which captures the date when the delay-causing incident occurred;
- route, which captures the number of the bus route;
- time, which captures when the delay-causing incident occurred;
- day, which captures day of the week;
- location, which captures the location of the delay-causing incident occurred;
- incident, which captures the description of the delay-causing incident;
- minimum delay, which captures the delay (in minutes) to the schedule for the following bus;
- minimum gap, which captures the total scheduled time (in minutes) from the bus ahead of the following bus (i.e., time gap between successive buses running the same route);
- direction, which captures the direction of the bus route; and
- vehicle, which captures the vehicle number.

The data are usually provided on a yearly basis—in an easily accessible format (e.g., Excel files)—and updated on a regular basis (e.g., multiple times per month). An example view of the data provided is in Table 1. Here, for the "direction" attribute, five possible values were expected:

- N/B, which indicates the northbound route;
- S/B, which indicates the southbound route;
- E/B, which indicates the eastbound route;
- W/B, which indicates the westbound route; and
- B/W, which indicates both ways.

Table 1. Sample bus delay data.

Report date	Route	Time	Day	Location	Incident	Min delay	Min gap	Direction	Vehicle
01-Jan-19	39	12:13:00 AM	Tuesday	NECR	Mechanical	9	18	W/B	1794
01-Jan-19	111	12:15:00 AM	Tuesday	Eglington	Mechanical	15	30	S/B	8065
01-Jan-19	35	12:18:00 AM	Tuesday	Finch	Mechanical	9	18	S/B	3275
01-Jan-19	25	12:30:00 AM	Tuesday	Don Mills Rd/ Eglinton Ave E	Mechanical	9	18	N/B	8840
01-Jan-19	36	12:40:00 AM	Tuesday	Humberwood	Investigation	9	18	E/B	9119
…	…	…	…	…	…	…	…	…	…

3.2 Data Cleaning

In the first step, our intelligent system removes NULL and incomprehensible values for each attribute/parameter name. Specific details are described below.

For the attribute/parameter "time" (i.e., time when the delay-causing incident occurred), our system converts the 12-hour time representation to 24-hour time representation. It then discretizes the data and bins the values into the following four equal-size intervals of 6 h each:

- Night, for 23:00–5:00 (i.e., 11:00PM–5:00 AM);
- Morning, for 5:00–11:00 (i.e., 5:00 AM–11:00 AM);
- Afternoon, for 11:00–17:00 (i.e., 11:00 AM–5:00 PM); and
- Evening, for 17:00–23:00 (i.e., 5:00 PM–11:00 PM).

Such a step is beneficial for the ease of frequent pattern mining and association rule mining as it helps us to get insights about (a) the frequency of delay intervals and (b) relationships among delay-causing incidents. Figure 1 shows the frequency of time and its corresponding discretized interval. The figure also reveals that most delays occur at 15:00, 8:00, and 14:00.

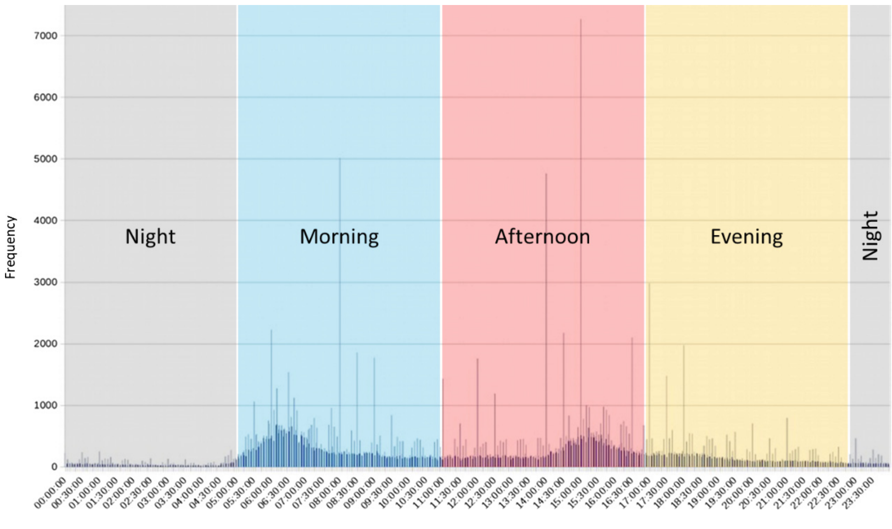

Fig. 1. Frequency of time of the day when the delay-causing incident occurred

For the attribute "location" (i.e., the location of the delay-causing incident occurred), it seems to be manually entered into the dataset. Consequently, this leads to inconsistencies in syntax—especially, in terms of spelling, special characters, and capitalization. Our system fixes this problem by imposing a consistent naming standard between all the location names in the database. It does so by utilizing related auxiliary datasets (say, data about routes and schedules[4]) that contain information for routes and stop locations—represented by bus stop names, as well as other identifiers (e.g., IDs, latitude and longitude coordinates, etc.) of bus stops. With standard stop names, our system conducts pairwise comparisons between any problematic bus stop name entry (i.e., any bus stop name that is unavailable from the dataset containing standard stop names) with standard ones. Specifically, the system applies fuzzy string matching to compare differences between the pairs of bus stop names and find the most similar match (of the stop name). As there may be more than a single similar bus stop name for a given problematic bus stop name entry, our system associates each potential match or similar bus stop name with a similarity value. The one with the highest similarity value is considered as the closest match for that problematic bus stop entry.

For the attribute "minimum gap" (i.e., the delay in minutes to the schedule for the following bus), our system discards insignificant delays (e.g., delays of less than 5 min). These very short delays are usually unavoidable but tolerable by bus riders. The remaining delays (i.e., those delayed for at least 5 min) are then discretized and binned into the following five categories, which indicate the delay severity:

- short delay of at least 5 min but less than 10 min,
- medium delay of at least 10 min but less than 20 min,
- long delay of at least 20 min but less than 30 min,
- severe delay of at least 30 min but less than 60 min, and
- crippling delay of at least 60 min.

We select these bins based data distribution (see Fig. 2) and general perception of bus delay times by transit riders.

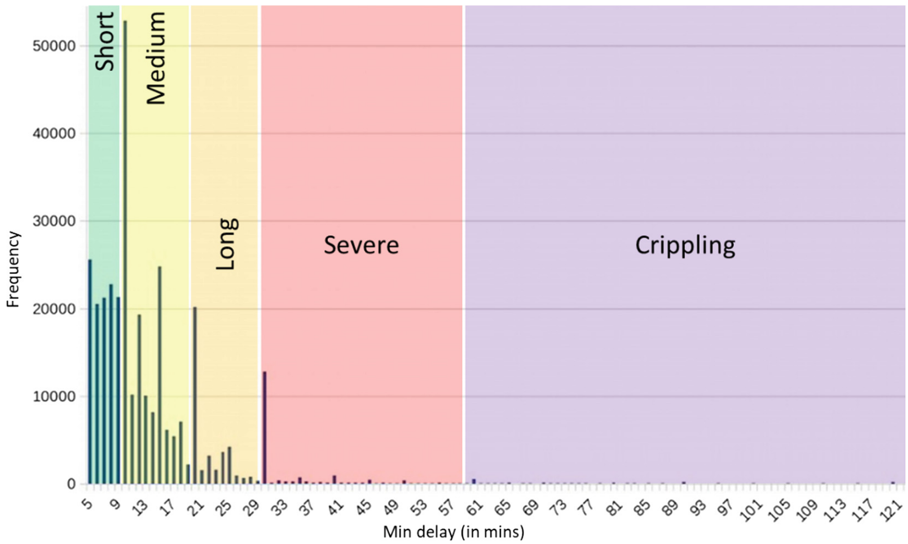

Fig. 2. Frequency of the minimum delay time

[4] https://portal0.cf.opendata.inter.sandbox-toronto.ca/dataset/ttc-routes-and-schedules/.

For the attribute "direction" (i.e., direction of the bus route), we know there supposes to be only five possible values (namely, N/B, S/B, E/B, W/B, and B/W). However, similar to the attribute "location", the direction also seems to be manually entered into the dataset. Consequently, this also leads to inconsistencies in syntax—especially, in terms of spelling and capitalization (e.g., "NB", "nb", "N"). Our system fixes these spelling and capitalization problems by finding the closet matches among the five possible values for direction. However, it can be challenging when the problematic direction entry shows "W". Does "W" mean westbound or both ways? Our system partially fixes this problem by again utilizing related auxiliary datasets (say, data about routes and schedules) that contain information for routes. Based on the routes, the system can determine that "W" means both ways if the route is a north-south one. However, if the route is an east-west one, the system may require additional information to precisely determine whether "W" means both ways or just westbound.

3.3 Data Analytics via Frequent Pattern Mining for Checking the Feasibility of Predictions

Once the data are cleaned, our intelligent system conducts data analytics to find evidence for supporting the conjecture that "patterns discovered from bus delay data in previous years are sufficient to predict bus delays in the future". To do so, we examine a few frequent pattern types, where the type of a frequent pattern is defined by the rows it includes, and count the occurrences of each type across years to get some insights about the data. For instance, we examine correlation between different attributes and severity of delay incurred. Moreover, we also examine the intersection of patterns between years and counted number of patterns by types. Frequent patterns are mined using frequent pattern mining algorithm such as FP-growth.

By using frequent pattern mining, we obtain the frequency of bus delays by (a) *time* of the day and (b) *day* of the week for the years 2014–2017. See Fig. 3.

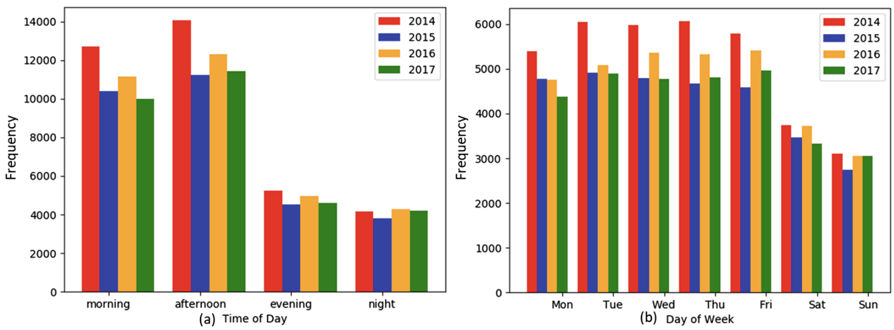

Fig. 3. Frequency of the (a) time of the day and (b) day of the week when the delay-causing incident occurred.

Observed from Fig. 3(a), bus delays throughout time of the day were consistent throughout all the years from 2014 to 2017. Moreover, bus delays were more frequent in the morning and afternoon than in the evening and at night. These patterns reveal that, as a majority of people commute to school/work in the morning, more congestion on the road, which leads to more bus delays. Similar scenarios for the afternoon when people come back from school/work in the afternoon.

Observed from Fig. 3(b), bus delays were most frequent in year 2014. Moreover, bus delays were more frequent during weekdays than weekends. These patterns reveal that people commute to work on weekdays, which results in higher probability of congestion on the roads, which leads to delayed buses.

As the patterns are consistent throughout all the years for bus delays by both (a) day of week and (b) time of the day, *patterns discovered from bus delay data in 2014–2016 are reasonably sufficient to predict bus delays in 2017.*

3.4 Feature Selection via Frequent Pattern Mining

Our intelligent system conducts frequent pattern mining on historical bus delay data to select feature for predicting future bus delays. Specifically, the system iterates through each year of bus delay data to mine frequent patterns in each of the historical years using FP-growth. Any attributes within our dataset can be taken into account when mining frequent patterns. Among the mined results, interesting results are obtained by using the following parameters:

- report date (or report month),
- route,
- time (or time interval),
- day,
- location (or standard bus stop name),
- incident, and
- minimum delay (or delay severity).

These attributes provide insight into a plethora of interesting delay factors, which we could further analyze to determine interesting results as to when, why, and where delays were frequently occurred.

Once these frequent patterns consisting of sets of parameters were computed, the frequent patterns were partitioned into two categories: (a) *previous* years and (b) a single *future* year. Next, we identify all of the frequently occurring patterns, which exist in every previous year. These were the patterns that occur most consistently between years, and they are good candidates—as *selected features*—for making accurate predictions for future year. See Table 2 for frequency of frequent patterns mined from (a–c) three previous years 2014–2016, (d) future year 2017, and (e–g) their three combinations of previous & future years within 2014–2017.

Table 2. Sample frequent patterns (i.e., selected features for predictive analytics).

Frequent patterns	Frequency						
	Previous years			Future year	Intersections of previous & future years		
	2014	2015	2016	2017	2016 ∩ 2017	2015 ∩ 2016 ∩ 2017	2014 ∩ 2015 ∩ 2016 ∩ 2017
1. {Route, time, delay severity}	200	160	170	152	128	101	97
2. {Route, delay severity}	174	148	161	149	138	125	123
3. {Route, direction, delay severity}	172	148	159	136	125	106	98
4. {Route, incident, delay severity}	160	134	147	120	104	83	77
5. {Month, incident, delay severity}	138	130	132	134	127	125	124
…	…	…	…	…	…	…	…
15. {month, delay severity}	52	52	54	53	51	48	45
16. {day, delay severity}	35	35	35	35	35	35	35
17. {incident, delay severity}	25	24	25	25	24	24	24
18. {Time, delay severity}	20	20	20	20	20	20	20

Observed from Table 2, simpler patterns consisting of two features (e.g., frequent patterns 15–18) are constantly frequent with (almost) the same frequency across individual years from 2014 to 2017. They are reliable features to be selected for predictive analytics. Moreover, more complex patterns consisting of more features (e.g., frequent patterns 1–5, each consist of three features) are also constantly frequent within a reasonable range of frequency across individual years from 2014 to 2017. They are also reliable features to be selected for predictive analytics.

As expected, frequency of the intersections does not increase when more years are added to the intersections. When frequency of the intersections of three previous years 2014–2016 is high, the patterns are very likely to occur in future year 2017 and thus are good predictive features. Along this direction, they would also be good predictive features in the true future (say, 2018, 2019, etc.) as well.

3.5 Predictive Analytics via Decision Tree Induction

Once the features are selected by frequent pattern mining, our intelligent system conducts predictive analytics. Specifically, it builds a decision tree, which stores the probability that a bus delay would occur given a user query with the following input parameters:

- a specific date including (a) month and (b) day of the week;
- time of the day; and
- bus stop location.

Our system stores the decision tree on disk, which gives us a viable solution that does not require computation/classification each time a query is made. For more accurate/time-relevant predictions, the system incorporates a time-fading model so that more recent bus delay data carry heavier weights than those older bus delay data. Also, whenever a new data entry comes in, the tree is updated (or a new tree is built) to reflect the new value. By doing so, the system would be able to provide real-time predictive functionality.

4 Validation of Our Intelligent System with Real-Life Open Data

To evaluate our intelligent system and validate its prediction results, we used the aforementioned TTC bus delay data available on the City of Toronto Open Data Portal.

Specifically, to evaluate the prediction accuracy of our system, once all of patterns are identified (i.e., features are selected), we examine the frequently occurring patterns of the future year, and create two lists:

- Frequent patterns occurring in all years, both previous and future (i.e., successful predictions); and
- Frequent patterns occurring in all previous years, but not the future year (i.e., unsuccessful predictions).

Table 3 shows some examples of successful and unsuccessful predictions. For instance, based on bus delay data in previous years 2014-2016, our intelligent system predicts that "mechanical problems occur on a Tuesday afternoon in January are likely to cause a short delay" in future year 2017, and appeared frequently in 2017. This is one of many successful predictions returned by our system.

Table 3. Sample predictions.

Frequent patterns (with feature values)	Delay severity	Previous years			Future year
		2014	2015	2016	2017
S1. {Jan, afternoon, Tue, mechanical}	Short	73	41	26	53
S2. {May, afternoon, Wed, mechanical}	Short	43	44	31	51
S3. {Mar, afternoon, Wed, mechanical}	Short	73	38	34	43
U1. {Mar, morning, Tue, mechanical}	Short	49	53	31	Infrequent
U2. {Mar, afternoon, Tue, mechanical}	Medium	49	36	32	Infrequent

In terms of prediction accuracy, our intelligent system is accurate. For instance, among those 1553 frequent patterns of length 4 (i.e., each frequent pattern consists of values for selected features/parameters like those shown in Table 3) returned by our system, 1,378 were successful predictions (i.e., 89% of the generated predictions are *true positives*) and only 175 were unsuccessful predictions (i.e., 11% of the generated predictions are *false positives*).

5 Conclusions

As we are living in the era of big data, big data are everywhere. With advances in technologies, huge volumes of a wide variety of valuable data—which may be of different levels of veracity—can be easily generated or collected at a high velocity. They can be originated from a wide variety of data sources in various real-life applications. Public transportation data or urban data are examples of big data.

Timely public transportation is important because it encourages more people to utilize transit and reduce the congestion in the roads. Delayed public transportation (e.g., delayed bus) has significant negative impacts on the quality of life for the general inhabitants in the city as people end up losing their precious time waiting for buses due to bus delay. In this paper, we present an intelligent predictive analytics system for transportation analytics on open data towards the development of a smart city. Specifically, our system analyzes *open data* (which are freely accessible by general public)—namely, the TTC bus delay data from the City of Toronto Open Portal—and apply a data mining task of *frequent pattern mining* (e.g., using the FP-growth algorithm) to show that historical bus delay data are relevant to making predictions for future bus delays. Our decision tree based model accurately predicts the severity of possible bus delays based on input parameters like the month, time, day, time, incident and/or location of the delay. In other words, our key contribution is our intelligent system for analyzing public transport data on bus arrival time and predicting on whether or not bus arrives on time or late.

Please note that, although most of the city transits have GPS (which makes real-time scheduling or prediction easier than using other historical data), these GPS data may be out of reach by general public. This explains why we focus on open data, which are freely accessible by general public, in this paper. However, as ongoing work, we are interested in examining how additional information (e.g., GPS data if available) would affect the prediction accuracy and the ability of real-time prediction.

Moreover, as real traffic delays often determine by many other external factors outside the normal traffic congestion (e.g., road accidents, events, road maintenance, temporary closure of routes, etc.). As a second direction for ongoing work, we are interested in incorporating other data sources (e.g., Google Maps, Waze)—for additional navigation information, real-life travel times, and route details—into our intelligent systems.

Furthermore, as a third direction for ongoing work, we are exploring future enhancements to our intelligent systems. For instance, we are exploring the use of random forests (rather than the decision tree) for predictive analytics.

In addition, as a fourth direction for ongoing work, we are also exploring *transfer learning* from predictive analytics on bus delay data to other public transit modes such as predictive analytics on streetcars and/or subways. Specifically, we are exploring the transfer of knowledge learned from the current paper on TTC bus delay data to delay data for other transportation modes such as streetcar and/or subway via transfer learning. Along the same direction, we are also exploring the transfer of knowledge to other jurisdictions and/or other transportation data via transfer learning.

Acknowledgments. The project is partially supported by (i) NSERC (Canada) and (ii) University of Manitoba.

References

1. Bellatreche, L., Leung, C., Xia, Y., Elbaz, D.: Advances in cloud and big data computing. Concurrency Comput. Prac. Exp. **31**(2), e5053:1–e5053:3 (2019)
2. Hoi, C.S.H., Khowaja, D., Leung, C.K.: Constrained frequent pattern mining from big data via crowdsourcing. In: BigDAS 2017. AISC, vol. 770, pp. 69–79 (2017)
3. Leung, C.K.: Big data analysis and mining. In: Advanced Methodologies and Technologies in Network Architecture, Mobile Computing, and Data Analytics, pp. 15–27 (2019)
4. Leung, C.K., Deng, D., Hoi, C.S.H., Lee, W.: Constrained big data mining in an edge computing environment. In: BigDAS 2017. AISC, vol. 770, pp. 61–68 (2017)
5. Leung, C.K., Hoi, C.S.H., Pazdor, A.G.M., Wodi, B.H., Cuzzocrea, A.: Privacy-preserving frequent pattern mining from big uncertain data. In: IEEE BigData 2018, pp. 5101–5110 (2018)
6. Abdalaal, R.M., Ho, C.N.M., Leung, C.K., Ohin, N.I., Ur-Rehman, S.H.: A remotely control dimming system for LED lamps with power factor correction. In: IEEE ECCE 2018, pp. 4721–4727 (2018)
7. Barkwell, K.E., Cuzzocrea, A., Leung, C.K., Ocran, A.A., Sanderson, J.M., Stewart, J.A., Wodi, B.H.: Big data visualisation and visual analytics for music data mining. In: IV 2018, pp. 235–240 (2018)
8. Camara, R.C., Cuzzocrea, A., Grasso, G.M., Leung, C.K., Powell, S.B., Souza, J., Tang, B.: Fuzzy logic-based data analytics on predicting the effect of hurricanes on the stock market. In: FUZZ-IEEE 2018, pp. 576–583 (2018)
9. Cox, T.S., Hoi, C.S.H., Leung, C.K., Marofke, C.R.: An accurate model for hurricane trajectory prediction. In: IEEE COMPSAC 2018, vol. 2, pp. 534–539 (2018)
10. Leung, C.K., Jiang, F., Souza, J.: Web page recommendation from sparse big web data. In: IEEE/WIC/ACM WI 2018, pp. 592–597 (2018)
11. Aoki, E., Otsuka, S., Ikenaga, T., Yatsuzuka, M., Tokiwa, K.: Study on regional transportation linkage system that enables efficient and safe movement utilizing LPWA. In: CISIS 2018. AISC, vol. 772, pp. 968–977 (2018)
12. Rajput, P., Toshniwal, D., Agggarwal, A.: Improving infrastructure for transportation systems using clustering. In: BDA 2017. LNCS, vol. 10721, pp. 129–143 (2017)
13. Kennedy, C., Steinberger, J., Gasson, B., Hansen, Y., Hillman, T., Havránek, M., Pataki, D., Phdungsilp, A., Ramaswami, A., Mendez, G.V.: Greenhouse gas emissions from global cities. Environ. Sci. Technol. **43**(19), 7297–7302 (2009)
14. Morency, C., Trépanier, M., Agard, B.: Measuring transit use variability with smart-card data. Transp. Policy **14**(3), 193–203 (2007)

15. Bañeres, D., Serra, M.: On the design of a system to predict student's success. In: CISIS 2018. AISC, vol. 772, pp. 274–286 (2018)
16. Chen, R., Tong, Y., Yang, J., Wu, M.: Residual reconstruction algorithm based on sub-pixel multi-hypothesis prediction for distributed compressive video sensing. In: CISIS 2018. AISC, vol. 772, pp. 599–605 (2018)
17. Khan, M., Javaid, N., Iqbal, M.N., Bilal, M., Zaidi, S.F.A., Raza, R.A.: Load prediction based on multivariate time series forecasting for energy consumption and behavioral analytics. In: CISIS 2018. AISC, vol. 772, pp. 305–316 (2018)
18. Kulla, E., Morita, S., Katayama, K., Barolli, L.: Route lifetime prediction method in VANET by using AODV Routing Protocol (AODV-LP). In: CISIS 2018. AISC, vol. 772, pp. 3–11 (2018)
19. Phankokkruad, M., Wacharawichanant, S.: Prediction of mechanical properties of polymer materials using extreme gradient boosting on high molecular weight polymers. In: CISIS 2018. AISC, vol. 772, pp. 375–385 (2018)
20. Sun, D., Luo, H., Fu, L., Liu, W., Liao, X., Zhao, M.: Predicting bus arrival time on the basis of global positioning system data. Transp. Res. Rec. **2034**(1), 62–72 (2007)
21. Lin, Y., Yang, X., Zou, N., Jia, L.: Real-time bus arrival time prediction: case study for Jinan. China J. Transport. Eng. **139**(11), 1133–1140 (2013)
22. Chien, S.I., Kuchipudi, C.M.: Dynamic travel time prediction with real-time and historic data. J. Transport. Eng. **129**(6), 608–616 (2003)
23. Kormáksson, M., Barbosa, L., Vieira, M.R., Zadrozny, B.: Bus travel time predictions using additive models. In: IEEE ICDM 2014, pp. 875–880 (2014)
24. Shalaby, A., Farhan, A.: Prediction model of bus arrival and departure times using AVL and APC data. J. Public Transport. **7**(1), 41–61 (2004)
25. Vanajakshi, L., Subramanian, S.C., Sivanandan, R.: Travel time prediction under heterogeneous traffic conditions using global positioning system data from buses. IET Intel. Transport. Syst. **3**(1), 1–9 (2009)
26. Williams, B.M., Hoel, L.A.: Modeling and forecasting vehicle traffic flow as a seasonal ARIMA process: theoretical basis and empirical results. J. Transport. Eng. **129**(6), 664–672 (2003)