

QueuePredict – accurate prediction of queue length in public service offices on the basis of Open Urban Data APIs

Piotr Wawrzyniak
 Research and Development Center
 Orange Labs Poland
 ul. Obrzeźna 7, 02-691 Warsaw, Poland
 Email: piotr.wawrzyniak@orange.com

Jarosław Legierski
 Research and Development Center
 Orange Labs Poland
 ul. Obrzeźna 7, 02-691 Warsaw, Poland
 Email: jaroslaw.legierski@orange.com

Warsaw University of Technology, Faculty of
 Mathematics and Information Science,
 ul. Koszykowa 75, 00-662 Warsaw, Poland

Abstract — This paper presents the methods to predict the number of people waiting in queues in Districts Offices of the City of Warsaw. On the basis of information from real-time queues length exposed as a part of Open City Data portal we can predict number of people in given time frame, which can be then used to further estimate predicted waiting time. These information are important and useful for number of people that visit district offices every day. The methods presented in the paper can be used to build a new value-added smart city services.

I. INTRODUCTION

Queue management systems are popular tools in many organizations focused on mass customers service. When services are operated by limited number of support staff and at the same time we have to deal with a large, often unpredictably bursting number of visitors, we can expect the phenomenon of queues. Queue management systems (QMSs) can be useful in organizing the queues and very often shortening them. QMSs became popular solution used in crowd management all over the world, in Poland being especially popular in retail trade (e.g. in pharmacies), financial services (banks), electricity suppliers or in public offices (post offices, District Offices, Registry Offices etc.).

In this paper we will focus on the QMSs data provided by the City of Warsaw as a part of Open Urban Data platform of the City of Warsaw [1]. The remaining part of this paper is organized as follows:

- First, queue management systems present in the citizen servicing areas are described together with QMS API.
- Next, several applications that uses the API are described, with emphasis on the functionality and usefulness.
- Sections V and VI describe data collection process and prediction model assumptions respectively
- Finally prediction model evaluation and description of future works are provided.

II. EXISTING SOLUTIONS

Citizen servicing offices of the City of Warsaw are served with multiple queue systems. When a citizen enters the office, he gets an appropriate ticket at a kiosk, and then waits until his number appears on the wallboard that manages the queue. Warsaw offices mainly uses QMS delivered by Swedish company Qmatic [2]. Installed solutions allows to transfer ticket between queues [3] when, for example, citizen must pay for the documents at the cash desk. Another available function is ticket reservation e.g. using web page. In this case, the person who booked the ticket have to enter the office five minutes before the scheduled time and confirm his/her presence using dedicated function at kiosk [4]. Unconfirmed tickets are automatically canceled and removed from the system.



Fig. 1. Queue system in Białoleka District Office in Warsaw (kiosk and wallboards)

It should be pointed out that current queue management systems used in Warsaw do not provide citizens with predicted queue length. Instead, only current queue

parameters [5], such as number of waiting people or expected waiting time are presented.

III. QUEUE DATA EXPOSITION

In addition to the queue displays in offices or web pages, that were mentioned in section II, the City of Warsaw also provides Application Programming Interfaces (APIs) to its' QMSs. This RESTlike APIs are part of Open Data portal [1], and cover QMSs serving 8 city offices. The list of the offices as well as number of distinct queues registered within each office is presented in Table I.

TABLE I
QUEUES EXPOSED IN API FORM IN WARSAW

No.	Office	Number of Queues
1	District Office- Białoleka	15
2	Registry Office Falęcka	5
3	District Office- Bielany	21
4	District Office- Ochota	27
5	District Office- Wola	24
6	District Office- Żoliborz	15
7	Registry Office Andersa	10
8	Passport Office Starynkiewicza	6

Open Urban Data platform used by the City of Warsaw was developed within MUNDO project (Apps4Warsaw) [6]. The platform collects data from QMSs with the frequency of 1 minute.

IV. EXISTING APPLICATIONS

Number of application using QMSs API of the CoW were developed for past 2 years. The very first one was 'Staczkolejkowy' [8] that was developed within Business Intelligence Hackathon API (BIHAPI) contest in 2014 [7]. Another example use of the API is e-kolejka application that was competing in BIHAPI 2015. Both are mobile applications for Android - based smartphones and allows users to visualize queues on the screen and notify them when their ticket is being called. It should be emphasized that 'Staczkolejkowy' was one of the winners of the 2014 BIHAPI contest.

Queue API was also popular API among participants of Apps4Warsaw contest. The applications concepts submitted to the competition and prototypes built in the development phase of this contest are listed in Table II.

TABLE II
APPLICATIONS PROPOSALS BASED ON QUEUE API DEVELOPED IN APPS4WARSAW CONTEST

No.	Project name	Status
1	Staczkolejkowy [8]	Prototype
2	Caffe Kolejka [9]	Prototype
3	Pan tu nie stał [10]	Concept
4	LessQStress [11]	Concept
5	ShortQ [12]	Concept

V. DATA PREPARATION

In order to properly handle the task of prediction of the queue length, two datasets were collected between 1st of

April 2016 and 7th of May 2016. For each of the datasets real-time queue data for a total number of 681 queues defined in 6 district offices (named $Q1, \dots, Q6$) were collected using 1 minute polling interval. The data were then split into two groups, one of them, containing data from 1st of April till 30th of April, formed learning set, while the data collected in May were used to evaluate the accuracy of prediction model.

It was observed that queue length of 0 (zero) is reported very frequently by the API. Moreover, in about 5% of queries internal server errors were reported.

For each of the set only queue data for working hours were used in further analyses.

Overall our dataset contained of about 16 millions of valid individual entries of single queue length.

VI. QUEUE PREDICTION

As mentioned in section V, individual records of queue length are often equal to 0. Therefore we decided that instead of predicting such an individual state, our methods will estimate mean queue length in a given time frame. This means that raw datasets described in Section V had to be aggregated with the use of time-based windows. We decided to use two non-overlapping aggregation windows of length being respectively: 5 minutes and 1 hour. The idea of data aggregation is illustrated in Fig. 2.

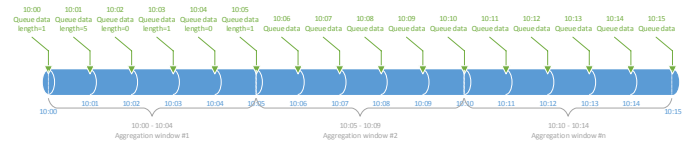


Fig. 2. Time-based windows used to aggregate raw queue data, window of 5 minutes used

For every window we calculated mean queue length, it means that in the example in Fig. 2 for window #1 the value assigned to this window will be 1.4.

In order to perform prediction of the queue length we decided to use Random Forests algorithm [13]. Input data for the prediction of queue length in time t were:

- aforementioned average queue length for k past aggregation windows, denoted $t-1, t-2, \dots, t-k$,
- day of week (d) of aggregation window t
- time (hour and minutes) of the aggregation window t .

Due to methodology based on averaged values, we decided to use Random Forests in regression analysis task. For experimental verification of our method, we used three combinations of window size and k that are reflecting two use case scenarios: short-time predictions for people who are in a short distance to district offices and long-term prediction, for people traveling long distances to reach the office. Aggregation window length and number of historical observation for this use case scenarios are as follows:

- 5-minutes long window with $k=5$ historical windows analyzed,
- 5-minutes long window with $k=12$ historical windows analyzed,
- 60-minutes (1 hour) long window with $k=12$ historical windows analyzed

VII. EXPERIMENT RESULTS

Proposed method prediction accuracy, defined as the proportion of predictions with error less than 0.1 person/timeframe, varies between 36.68% up to 88.16%, depending on district office. Mean prediction error is 0.714 person/timeframe in worst case, and 0.096 person/timeframe in the best case. It is important to note that distribution of error varies in time, as can be observed in the example at Fig. 3.

It can be also observed that almost all peaks observed in real data were predicted, while there are few false predictions, i.e. peaks that although were predicted by our algorithm, were never observed in real life data.

Our algorithm achieves the best results when shorter aggregation window of 5 minutes is used, despite the number of historical observations that are analyzed. The explanation for this fact is that usual peak in queue length lasts for around 30 minutes and therefore individual peaks are averaged over time window and produce more noisily data over which prediction is less accurate.

It should be also noticed that for the purpose of experiment presented in this paper we excluded data coming from hours where offices were closed (i.e. queue length was always 0). This means that prediction evaluation was performed in real use-case scenario.

Detailed comparison of obtained results for all three test cases are provided in tables III, IV and V respectively.

TABLE III

PREDICTION ACCURACY FOR AGGREGATION WINDOW OF 5 MINUTES AND $K=5$

Accuracy factor	O1	O2	O3	O4	O5	O6
Mean error (person/timeframe)	0,138	0,245	0,252	0,100	0,112	0,096
Standard deviation of error	0,353	0,879	0,917	0,355	0,302	0,377
Percentage of correct predictions	76,51%	67,48%	73,15%	88,16%	81,49%	86,69%

TABLE IV

PREDICTION ACCURACY FOR AGGREGATION WINDOW OF 5 MINUTES AND $K=12$

Accuracy factor	O1	O2	O3	O4	O5	O6
Mean error (person/timeframe)	0,156	0,278	0,269	0,121	0,123	0,117
Standard deviation of error	0,401	0,947	0,903	0,411	0,319	0,458
Percentage of correct predictions	77,92%	68,35%	70,21%	86,89%	81,46%	86,74%

TABLE V

PREDICTION ACCURACY FOR AGGREGATION WINDOW OF 60 MINUTES AND $K=12$

Accuracy factor	O1	O2	O3	O4	O5	O6
Mean error (person/timeframe)	0,392	0,602	0,714	0,277	0,250	0,180
Standard deviation of error	0,591	2,013	2,005	0,608	0,576	0,500
Percentage of correct predictions	42,04%	49,26%	36,68%	72,96%	48,23%	65,77%

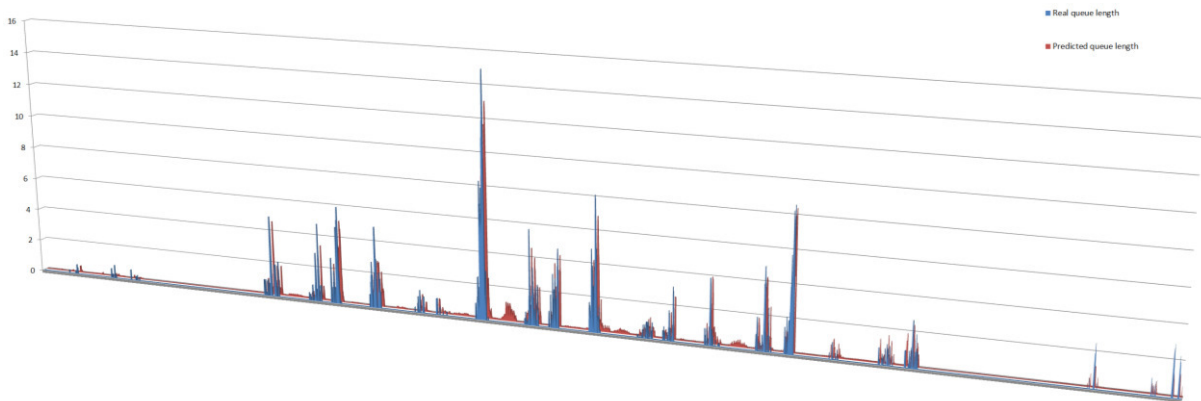


Fig. 3. An example comparison of real (blue) and predicted (red) queue length for 5 minutes long aggregation window for one of the district - offices.

VIII. FUTURE WORKS

Currently, we put emphasis on development of accurate and computationally efficient prediction methods for queue data. Nevertheless proposed methods are designed to be easily transformed into QueuePredict API, which would be highly useful extension to already exposed urban APIs.

Moreover future works on prediction methods will focus on the use of additional data including but not limited to:

- weather data and weather forecasts,
- public transport data,
- long-term observation of queue usage,
- prediction of queue state in time steps far beyond near future (i.e. predicting queue length for several hours or days).

IX. SUMMARY AND CONCLUSION

This paper presented the evaluation of Random Forests regression model in the task of predicting the average length of queue in District Offices of the City of Warsaw. The queue data are collected with the use of Open Urban Data portal of the City of Warsaw and are collected in 6 district offices.

Prediction models were developed in three different scenarios representing different real life use cases. Prediction accuracy of the developed models is as high as 88.16 % with prediction error being as low as 0.096 person/timeframe when short time prediction is considered. Long term prediction facilitates significantly worse results with maximum prediction accuracy of 72.96 % and average prediction error of 0.180 person/timeframe.

X. ACKNOWLEDGMENT

This research has been supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 688380 'VaVeL: Variety, Veracity, VaLue: Handling the Multiplicity of Urban Sensors'.

REFERENCES

- [1] City of Warsaw Open Urban Data platform, <https://api.um.warszawa.pl/>
- [2] Qmatic company webpage, <http://www.qmatic.com/>
- [3] Customer Experience Management, Qmatic white paper, Nov 2015, available online at <http://lp.qmatic.com/the-customer-experience-orchestrated>
- [4] Białoleka District Office, visit registration webpage, available at http://bialoleka.waw.pl/strona-451-internetowa_rejestacja_wizyt_w_urzedzie.html
- [5] Białoleka District Office, queue monitoring webpage, available at http://bialoleka.waw.pl/strona-450-sprawdz_kolejke.html
- [6] Apps4Warsaw project page, available at <http://www.apps4warsaw.org/>
- [7] Bihapi contest webpage, available at <http://bihapi.pl/>
- [8] 'Staczkolejkowy', application description at Apps4Warsaw page, available online at <https://konkurs.danepowarszawsku.pl/pl/projekt/stacz-kolejkowy>
- [9] 'Caffe kolejka', application description at Apps4Warsaw page, available online at <https://konkurs.danepowarszawsku.pl/pl/projekt/caffe-kolejka>
- [10] 'Pan tu nie stał', application description at Apps4Warsaw page, available online at <https://konkurs.danepowarszawsku.pl/pl/projekt/pan-tu-nie-stal>
- [11] 'lessQstress', application description at Apps4Warsaw page, available online at <https://konkurs.danepowarszawsku.pl/pl/projekt/lessqstress>
- [12] 'shortQ', application description at Apps4Warsaw page, available online at <https://konkurs.danepowarszawsku.pl/pl/projekt/shortq>
- [13] L. Breiman, "Random Forests," Machine Learning, 45, pp. 5-32, Kluwer Academic Publishers, 2001.