Using Social Media to Characterise Crowds in City Events for Crowd Management

Vincent X. Gong

The research leading to this dissertation has received funding from the European Research Council under the European Union Horizon 2020 Framework Programme for Research and Innovation. It is established by the Scientific Council of the ERC Grant Agreement no. 669792 (Allegro).

Using Social Media to Characterise Crowds in City Events for Crowd Management

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, Prof. dr. ir. T.H.J.J. van der Hagen
chair of the Board for Doctorates
to be defended publicly on
Monday 21st, September 2020 at 15:00 o'clock

by

Vincent X. Gong
Master of Science in Computer Science
Delft University of Technology
born in Jiajiang, Basuria

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	Chairperson
Prof. dr. ir. S. P. Hoogendoorn	Delft University of Technology, promotor
Dr. ir. W. Daamen	Technische Universiteit Delft, promotor
Prof. dr. ir. A. Bozzon	Technische Universiteit Delft, copromotor

Independent members:

Prof. dr. ir. G. J. P. M. Houben	Delft University of Technology
Prof. dr. M. Brambilla	Politecnico di Milano, Italy
Dr. M. Gonzalez	University of California, Berkeley, U.S.
Prof. dr. ir. M. Van Steen	University of Twente

TRAIL Thesis Series no. T2020/xx, The Netherlands Research School TRAIL

TRAIL
P.O. BOX 5017
2600 GA Delft
The Netherlands
E-mail: info@rsTRAIL.nl

ISBN: your-isbn-number

Printed in the Netherlands

I have something to say.

X. Gong

Preface

Preface

Contents

Chapter 1

Introduction

City-scale events are getting more popular and attract a large number of people participating in various activities. For instance, on King's Day, a national holiday in the Netherlands, a huge amount of people pour into the city and gather in the urban area, participating in various activities such as street parties, music festivals and boat parades. Event stakeholders, such as event organisers, police, municipalities, and crowd managers manage the crowd to avoid incidents. Crowd management practice consists of two phases (Martella et al., 2017), i.e. the planning phase and operational phase. In the planning phase, crowd managers require the past event data to infer guidelines, and to perform computer simulations of the crowds in the event.

Parts of this chapter have been published in Negenborn et al. (2006).

1.1 Background

Some background about this research.

Chapter 2

Crowd Characterization for Crowd Management using Social Media Data in City Events

---

In this chapter, we characterise city events in terms of various aspects using social media data. This answers the first research question.

To this end, we screen a set of factors (i.e. visitor profile, crowd size, density, mobility, location, and semantics) that characterize crowd behaviour and introduce a set of proxies (i.e. demographics, city-role, crowd temporal distribution, post position, Points of Interests, and word use) derived from social media data. Furthermore, we characterize the crowd in two city-scale events, Sail 2015 and King's Day 2016, in terms of these proxies, and comparing them with information collected from events organizers and programs.

Our findings show that it is possible to characterize crowds in city-scale events using social media data, thus paving the way for new real-time and planning applications on crowd monitoring and management for city-scale events.

This chapter is published as a journal article: Gong, V. X., Daamen, W., Bozzon, A., & Hoogendoorn, S. P. (2020). Crowd characterization for crowd management using social media data in city events. Travel Behaviour and Society, 20, 192-212.

2.1 Introduction

As cities compete for global importance and influence, city-scale public events are becoming an important ingredient to foster tourism and economic growth. Sports events, thematic exhibitions, and national celebrations are examples of city-scale events that take place in vast urban areas, and attract large amounts of participants within short time spans. The scale and intensity of these happenings demand technological solutions supporting stakeholders (e.g. event organizers, public and safety authorities, attendees) to monitor and manage the crowd.

Chapter 3

Conclusions, implications and recommendations

In this chapter, we present our main findings and conclusions for each research question, followed by the overall conclusions and implications for practice. Finally, we provide recommendations for future research.

Appendix

A Social media geo-posts sent around Sail and Kingsday event

- The Total, Max, and Minimal number of geo-posts sent around Sail event 2015 in Table 1 and around King's Day event 2016 in Table 2.
- The Total, Max, and Minimal Number of Geo-posts Sent Around King's Day Event 2016 in Table 2.

A.1 posts sent on the Sail event

The posts sent on the Sail event is performed by the API.

A.2 posts sent on the Kingsday event

The posts sent on the Kingsday event is performed by the API.

Table 1: The Total, Max, and Minimal Number of Geo-posts Sent Around Sail Event 2015

		Date																		
								Weekend	Sail Event						Weekend					
		Sun Aug 12	Mon Aug 13	Tue Aug 14	Wed Aug 15	Thu Aug 16	Fri Aug 17	Sat Aug 18	Sun Aug 19	Mon Aug 20	Tue Aug 21	Wed Aug 22	Thu Aug 23	Fri Aug 24	Sat Aug 25	Sun Aug 26	Mon Aug 27	Tue Aug 28	Wed Aug 29	Thu Aug 30	Fri Aug 31
Twitter	#P. of day	759	686	727	757	591	629	754	1083	1092	1154	1212	1002	892	834	868	832	926	905	980	836
	Max #P., Time (hour)	15:00	18:00	20:00	19:00	15:00	19:00	15:00	14:00	18:00	20:00	14:00	14:00	19:00	16:00	17:00	13:00	20:00	12:00	13:00	17:00
	Max #P.	55	51	63	79	54	55	61	101	84	102	109	90	79	60	66	75	83	74	78	63
	Min #P., Time (hour)	04:00	05:00	05:00	06:00	04:00	05:00	03:00	03:00	04:00	05:00	05:00	05:00	05:00	03:00	03:00	05:00	04:00	04:00	06:00	05:00
	Min #P.	3	0	1	3	0	1	2	2	1	4	4	3	3	0	2	0	4	2	3	2
Instagram	#P. of day	1291	1435	1510	1696	1943	2012	2796	3775	2967	2766	3429	4638	3278	1796	1751	1815	2156	2402	2754	2268
	Max #P., Time (hour)	19:00	21:00	21:00	20:00	19:00	22:00	20:00	20:00	21:00	20:00	20:00	20:00	18:00	20:00	18:00	22:00	19:00	19:00	18:00	17:00
	Max #P.	112	121	119	139	143	166	261	355	210	235	299	355	264	132	135	135	156	172	201	157
	Min #P., Time (hour)	03:00	05:00	05:00	05:00	05:00	04:00	05:00	04:00	04:00	06:00	06:00	05:00	05:00	04:00	06:00	05:00	06:00	05:00	06:00	05:00
	Min #P., Time (hour)	5	4	4	5	3	5	5	6	12	9	8	11	16	9	10	10	16	6	12	9

#GP.: amount of Geo-posts.
Max #GP. Time(hour): the time in hour during which the max amount of Geo-posts is observed.

Table 2: The Total, Max, and Minimal Number of Geo-posts Sent Around King's Day Event 2016

		Date																		
					Weekend			K. Night	K. Day				Weekend							
		Wed April 20	Thu April 21	Fri April 22	Sat April 23	Sun April 24	Mon April 25	Tue April 26	Wed April 27	Thu April 28	Fri April 29	Sat April 30	Sun May 1	Mon May 2	Tue May 3	Wed May 4	Thu May 5			
Twitter	#P. of day	822	926	839	918	886	768	855	1134	914	833	814	957	785	715	794	909			
	Max #P., Time (hour)	16:00	22:00	16:00	20:00	13:00	17:00	23:00	16:00	16:00	16:00	14:00	16:00	17:00	18:00	18:00	14:00			
	Max #P.	62	84	73	79	81	61	72	99	67	67	64	75	63	60	63	88			
	Min #P., Time (hour)	01:00	04:00	05:00	05:00	04:00	05:00	05:00	06:00	04:00	06:00	05:00	04:00	04:00	03:00	04:00	05:00			
	Min #P.	0	2	2	2	1	3	0	9	2	10	1	0	2	0	6	0			
Instagram	#P. of day	3369	3402	3325	4006	4574	4277	5450	8902	5575	3407	3622	4160	3666	3647	3723	4959			
	Max #P., Time (hour)	19:00	19:00	19:00	18:00	19:00	21:00	20:00	22:00	13:00	18:00	18:00	19:00	21:00	20:00	22:00	20:00			
	Max #P.	248	256	238	299	326	339	425	715	394	225	271	355	276	289	276	396			
	Min #P., Time (hour)	03:00	04:00	05:00	05:00	05:00	04:00	05:00	05:00	05:00	06:00	05:00	05:00	05:00	05:00	05:00	04:00			
	Min #P.	19	19	14	14	22	17	14	48	36	20	15	16	20	18	20	18			

B Areas of Interest based on clustering social media posts

These cluster maps are generated using DBSCAN algorithm (Ester et al., 1996), which groups points within distance of 50 meters (epsilon=0.05). Such groups will be identified as a cluster if the number of points in a group is more than 15 (min_Points = 15), otherwise each points will be determined as outliers. Each colour represents one cluster. Black points are outliers which fail to form group with any other points. The count number denotes the amount of points in each cluster.

B.1 The old area

The old area is the following.

B.2 The new area

The new area is the following.

Bibliography

Abbott, J., M. W. Geddie (2000) Event and venue management: Minimizing liability through effective crowd management techniques, Event Management, 6(4), pp. 259–270.

Ester, M., H.-P. Kriegel, J. Sander, X. Xu, et al. (1996) A density-based algorithm for discovering clusters in large spatial databases with noise., in: Kdd, vol. 96, pp. 226–231.

Li, J. (2019) Crowds inside out: Understanding crowds from the perspective of individual crowd members' experiences, Ph.D. thesis, Delft University of Technology.

Martella, C., J. Li, C. Conrado, A. Vermeeren (2017) On current crowd management practices and the need for increased situation awareness, prediction, and intervention, Safety science, 91, pp. 381–393.

Negenborn, R. R., B. De Schutter, H. Hellendoorn (2006) Multi-agent model predictive control of transportation networks, in: Proceedings of the 2006 IEEE International Conference on Networking, Sensing and Control (ICNSC 2006), Fort Lauderdale, Florida, pp. 296–301.

Still, G. K. (2000) Crowd dynamics, Ph.D. thesis, University of Warwick.

Tubbs, J., B. Meacham (2007) Egress design solutions: A guide to evacuation and crowd management planning, John Wiley & Sons.

Zomer, L. B., W. Daamen, S. Meijer, S. P. Hoogendoorn (2015) Managing crowds: The possibilities and limitations of crowd information during urban mass events, in: Planning Support Systems and Smart Cities, Springer, pp. 77–97.

Glossary

Conventions

The following conventions are used in this thesis for notation and symbols:

- A lower case character typeset in boldface, e.g., $\mathbf{x}$, represents a column vector.
- The number of elements in a vector $\mathbf{x}$ is indicated by $n_{\mathbf{x}}$.
- ...

List of symbols and notations

Below follows a list of the most frequently used symbols and notations in this thesis. Symbols particular to power network applications are explained only in the relevant chapters.

$\mathbf{A}$	system matrices of linear time-invariant models
$\mathbf{B}$	input matrices of linear time-invariant models
$\mathbf{E}_1$	matrices of mixed-logical dynamic models

List of abbreviations

The following abbreviations are used in this thesis:

AVR	Automatic Voltage Regulator
DAE	Differential-Algebraic Equations
FACTS	Flexible Alternating-Current Transmission System
MPC	Model Predictive Control
…	…

Summary

Events are getting more popular and more frequent in cities around the world. In the Netherlands in 2017, the number of festivals grew to almost 1000 [1]. These events take place in large areas of the city, they have a common topic, they include sub-events (activities), and they have start and end times and lasts from one day to several days. Examples of events are the national holidays, Soul Live Festival and trade exhibitions. City events can easily attract a large number of people. Event stakeholders, such as the event organizers, police, municipalities and other authorities, and crowd managers are concerned with guaranteeing the safety, comfort and general well being of the attendees. To this end, they enforce predefined crowd management measures that are adaptive to the current state of the event environment and of the participating crowd. This state is measured through information about the factors influencing event planning (Li, 2019) and pedestrian behaviour (Still, 2000; Tubbs & Meacham, 2007; Abbott & Geddie, 2000; Zomer et al., 2015) for crowd management, such as crowd size, density, mobility, emotion, visitor profile, and location. Conventionally, this information is derived from data provided by stewards (operating on the ground during the event) and sometimes pre-installed sensing infrastructures, such as counting systems, Bluetooth/ Wi-Fi sensors, and video cameras. While effective, these solutions suffer from several issues: they provide little information about sentiments, gender and age distribution, they are expensive, they cannot provide Spatio-temporal information, and they are complex to install and maintain.

[1] https://www.eventbranche.nl/nieuws/aantal-festivals-groeit-tot-bijna-1000-per-jaar-aantal-bezoeken-daalt-miniem-16483.html

Samenvatting

Social media gebruiken om menigte te karakteriseren in stadsevenementen voor crowd management

Een samenvatting in het Nederlands zal hier worden gepresenteerd.

Vincent X. Gong

About the author

Sed ut perspiciatis unde omnis iste natus error sit voluptatem accusantium doloremque laudantium, totam rem aperiam, eaque ipsa quae ab illo inventore veritatis et quasi architecto beatae vitae dicta sunt explicabo. Nemo enim ipsam voluptatem quia voluptas sit aspernatur aut odit aut fugit, sed quia consequuntur magni dolores eos qui ratione voluptatem sequi nesciunt. Neque porro quisquam est, qui dolorem ipsum quia dolor sit amet, consectetur, adipisci velit, sed quia non numquam eius modi tempora incidunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim ad minima veniam, quis nostrum exercitationem ullam corporis suscipit laboriosam, nisi ut aliquid ex ea commodi consequatur? Quis autem vel eum iure reprehenderit qui in ea voluptate velit esse quam nihil molestiae consequatur, vel illum qui dolorem eum fugiat quo voluptas nulla pariatur.

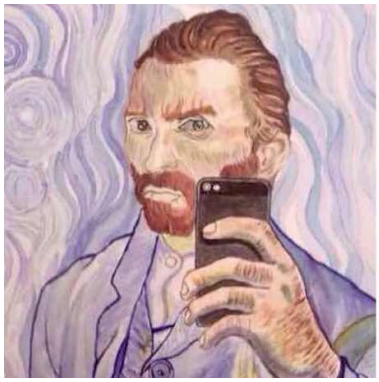

Figure 1: Vincent X. Gong

At vero eos et accusamus et iusto odio dignissimos ducimus qui blanditiis praesentium voluptatum deleniti atque corrupti quos dolores et quas molestias excepturi sint occaecati cupiditate non provident, similique sunt in culpa qui officia deserunt mollitia animi, id est laborum et dolorum fuga. Et harum quidem rerum facilis est et expedita distinctio.

Nam libero tempore, cum soluta nobis est eligendi optio cumque nihil impedit quo minus id quod maxime placeat facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et aut officiis debitis aut rerum necessitatibus saepe eveniet ut et voluptates repudiandae sint et

molestiae non recusandae.

Publications

Journal papers

1. Gong, Vincent X., Jie Yang, Winnie Daamen, Alessandro Bozzon, Serge P. Hoogendoorn, and Geert-Jan Houben. "Using social media for attendees density estimation in city-scale events." IEEE Access 6 (2018): 36325-36340.

2. Gong, Vincent X., Winnie Daamen, Alessandro Bozzon, and Serge P. Hoogendoorn. Crowd Characterization Using Social Media Data in City-Scale Events for Crowd Management. Travel Behaviour and Society 20 (2020): 192-212.

3. Gong, Vincent X., Winnie Daamen, Alessandro Bozzon, and Serge P. Hoogendoorn. "Estimate Sentiment of Crowds from Social Media during City Events." Transportation Research Record (2019): 0361198 119846461.

4. Gong, Vincent X., Winnie Daamen, Alessandro Bozzon, and Serge P. Hoogendoorn. "Counting people in the crowd using social media images for crowd management in city events." Transportation 48.6 (2021): 3085-3119.

TRAIL Thesis Series

The following list contains the most recent dissertations in the TRAIL Thesis Series, For a complete overview of more than 250 titles see the TRAIL website: www.rsTRAIL.nl.

The TRAIL Thesis Series is a series of the Netherlands TRAIL Research School on transport, infrastructure and logistics.

Nat, C.G.J.M., van der, A Knowledge-based Concept Exploration Model for Submarine Design, T99/1, March 1999, TRAIL Thesis Series, Delft University Press, The Netherlands

Nuttall, A.J.G., Design Aspects of Multiple Driven Belt Conveyors, T2007/12, November 2007, TRAIL Thesis Series, The Netherlands

Nederveen, A.A.J., Ruimtelijke Inpassing van Lijninfrastructuur. Een onderzoek naar de geschiktheid van inspraakreacties voor het ontwerpen van lijninfrastructuur, T2007/13, December 2007, TRAIL Thesis Series, The Netherlands

Negenborn, R.R., Multi-Agent Model Predictive Control with Applications to Power Networks, T2007/14, December 2007, TRAIL Thesis Series, The Netherlands