CETRA 2014

3rd International Conference on Road and Rail Infrastructure
28–30 April 2014, Split, Croatia

Road and Rail Infrastructure III

Stjepan Lakušić – EDITOR

Organizer
University of Zagreb
Faculty of Civil Engineering
Department of Transportation

CETRA 2014
3rd International Conference on Road and Rail Infrastructure
28–30 April 2014, Split, Croatia

TITLE
Road and Rail Infrastructure III, Proceedings of the Conference CETRA 2014

EDITED BY
Stjepan Lakušić

ISSN
1848-9850

PUBLISHED BY
Department of Transportation
Faculty of Civil Engineering
University of Zagreb
Kačićeva 26, 10000 Zagreb, Croatia

DESIGN, LAYOUT & COVER PAGE
minimum d.o.o.
Marko Uremović · Matej Korlaet

PRINTED IN ZAGREB, CROATIA BY
"Tiskara Zelina", April 2014

COPIES
400

Zagreb, April 2014.

Proceedings of the
3rd International Conference on Road and Rail Infrastructures – CETRA 2014
28–30 April 2014, Split, Croatia

Road and Rail Infrastructure III

EDITOR
Stjepan Lakušić
Department of Transportation
Faculty of Civil Engineering
University of Zagreb
Zagreb, Croatia

CETRA 2014
3rd International Conference on Road and Rail Infrastructure
28–30 April 2014, Split, Croatia

ORGANISATION

CHAIRMEN

Prof. Stjepan Lakušić, University of Zagreb, Faculty of Civil Engineering
Prof. Željko Korlaet, University of Zagreb, Faculty of Civil Engineering

ORGANIZING COMMITTEE

Prof. Stjepan Lakušić
Prof. Željko Korlaet
Prof. Vesna Dragčević
Prof. Tatjana Rukavina
Assist. Prof. Ivica Stančerić
dr. Maja Ahac
Ivo Haladin
dr. Saša Ahac
Josipa Domitrović
Tamara Džambas

All members of CETRA 2014 Conference Organizing Committee are professors and assistants of the Department of Transportation, Faculty of Civil Engineering at University of Zagreb.

INTERNATIONAL ACADEMIC SCIENTIFIC COMMITTEE

Prof. Vesna Dragčević, University of Zagreb
Prof. Isfendiyar Egeli, Izmir Institute of Technology
Prof. Rudolf Eger, RheinMain University
Prof. Ešref Gačanin, Univeristy of Sarajevo
Prof. Nenad Gucunski, Rutgers University
Prof. Libor Izvolt, University of Zilina
Prof. Lajos Kisgyörgy, Budapest University of Technology and Economics
Prof. Željko Korlaet, University of Zagreb
Prof. Zoran Krakutovski, University of Skopje
Prof. Stjepan Lakušić, University of Zagreb
Prof. Dirk Lauwers, Ghent University
Prof. Zili Li, Delft University of Technology
Prof. Janusz Madejski, Silesian University of Technology
Prof. Goran Mladenović, University of Belgrade
Prof. Otto Plašek, Brno University of Technology
Prof. Vassilios A. Profillidis, Democritus University of Thrace
Prof. Carmen Racanel, Technical University of Civil Engineering Bucharest
Prof. Tatjana Rukavina, University of Zagreb
Prof. Andreas Schoebel, Vienna University of Technology
Prof. Mirjana Tomičić-Torlaković, University of Belgrade
Prof. Audrius Vaitkus, Vilnius Gediminas Technical University
Prof. Nencho Nenov, University of Transport in Sofia
Prof. Marijan Žura, University of Ljubljana

CETRA 2014 28–30 April 2014, Split, Croatia
3rd International Conference on Road and Rail Infrastructure

STUDY ON THE AVAILABILITY OF "TWITTER" DATA FOR FORECASTING SUSPENSION TIME OF RAILWAY OPERATION

Makoto Fujiu[1], **Kazuyuki Takada**[2]
1 Kanazawa University, Department of Civil Engineering, Japan
2 Tokyo Denki University, Department of Civil Engineering, Japan

Abstract

Recently, a lot of people use Social Networking Service (SNS) to report their situation and do some findings around them. Especially, "twitter" is one of the easy ways to report some information for followers, and "twitter" users could send information and personal situations around them to "twitter" platform in any time. The amount of "twitter" data might be a lot more than the existing statistical data since some information are being uploaded at any time the users find it. Therefore, the numbers and the contents of "tweets" are different depending on the scale of the event, and a relevant key word is reported to "twitter" by the users. For example, when the railway operation stops, "twitter" users report that situation for the followers. The contents of the "tweets" are railway accident information, recovery time and complaint to railway company/staffs and so on. Then the duration of suspension time of the railway operation can be forecasted using the "twitter" data of the railway users. In this study, number and contents of "tweet" are analysed under accidental conditions which are a small railway accident and a railway accident in which a lot of people get influenced. Furthermore, the authors have developed suspension time forecasting model using "twitter" data, and the forecasted time is verified by comparing with the existing statistical data which is utilized by the Ministry of Land, Infrastructure, Transport and Tourism in Japan.

Keywords: railway, Social Networking Service, twitter, forecasting model

1 Introduction

There are many causes to delay the railway service schedule. And many passengers are assumed to be suffering a loss due to these delays. To investigate the current condition of the delay in Tokyo Metropolitan Area (TMA), statistics by Ministry of Land, Infrastructure, Transport and Tourism (MLIT) can be utilized. This statistics compiles the reports regarding railway accident and trouble with more than 30 minutes delay. According to this statistics, we can know when the accident happened, what the cause was, how large the influence was, and etc... According to the statistics that the transition of the number of railway accidents. It had been decreasing until 2003. The reason is that MLIT devised measures to reduce the number of railway accidents. For example, MLIT indicated railway companies to install emergency stop buttons in platforms, fall detection mats into railroads, and evacuation spaces under platforms [1],[2].
Recently, a lot of people use Social Networking Service (SNS) to report their situation and do some findings around them. Especially, "twitter" is one of the easy ways to report some information for followers, and "twitter" users could send information and personal situations around them to "twitter" platform in any time. For example, when the railway operation stops, "twitter" users report that situation for the followers. The contents of the "tweets" are railway

accident information, recovery time and complaint to railway company/staffs and so on. Then the duration of suspension time of the railway operation can be forecasted using the "twitter" data of the railway users. In this study, number and contents of "tweet" are analysed under accidental conditions which are a small railway accident and a railway accident in which a lot of people get influenced. Furthermore, the authors have developed suspension time forecasting model using "twitter".

2 DATA

2.1 Tweet Data Collection

Table 1 shows the twitter data which are created by NTT DATA Corporation. The data has created date and tweet text data in Japanese. Authors get tweet data regarding railway accident using key words which are "delay", "stop", "situation" and so on. Conditions of accident selection are date, time, scale of accident and modality of accident. Ten railway accidents were selected for this analysis as shown in table 2. An average suspension time is about 53.9 (minutes).

Table 1 "Twitter" data (in Japanese)

created_at(JST)	tweet text data
2012/8/3 16:05	強化練3日目終了です！U･x･U 今朝は東上線の遅れに巻き込まれた部員もいましたが、無事活動することができました。土日は休みなので、ちゃんと休んで来週の強化練に備えましょう！
2012/8/3 16:03	@hanon_kazuki そうかい？ RT 東上線人身事故。運転見合わせ中。朝から困ったもんだ。
2012/8/3 15:59	「人が死にやすい鉄道」1位は東武東上線、死亡率でトップだそうです。確かによく止まっているから。http://t.co/XUdEySvI来年から、一緒になる東急東横線みなとみらい線も二次被害？マジで最悪だね。
2012/8/3 15:36	@dancebook1 東上線は1時間半ほど待たされました(^^;;最近また事故か多いですね。
2012/8/3 15:34	@makadamia123 東上線、人身事故多いよね！(>_<)
2012/8/3 15:22	おいおい東上線人身事故とか朝からワイルドだな
2012/8/3 15:20	東上線運行情報 現在、事故・遅延に関する情報はありません。12/08/03 15:18 #駅の伝言板 #東上線
2012/8/3 15:14	そういえば、朝、東武東上線止まってたな…
2012/8/3 15:13	朝の路線別事故が気になったので考察。多いのは東上線、宇都宮線、池袋線などで埼玉に絡む路線から北上する路線に多い。ところが京急や田園都市線なんかは話にも上がらない。
2012/8/3 15:12	今日は朝から会社の説明会に行ったのですが志木駅に着くと東上線が止まっていました、絶望ですね、初めて行く説明会遅刻ですよ、幸先悪いですね、(続く)
2012/8/3 15:05	RT @Sociometry_487: え、東上線ダイヤ乱れたんすかw
2012/8/3 15:04	今日は東武東上線がやられたのか。昨日やられてたら研修死んでたわー(^^;;
2012/8/3 15:03	え、東上線ダイヤ乱れたんすかw
2012/8/3 15:00	15時http://t.co/Yz2X3vcF ブラックサンダー,武蔵野線,會澤翼,東武東上線,バドミントン,大原麗子,東武東上線 運行状況,マルセル・グエン,無気力試合,身長偏差値,梅ちゃん先生,フアンマルティン・デルポトロ,宇佐美貴史,スイムオフ,榮倉奈々,錦織圭,北島

Table 2 Selected railway accidents

Case	Year	Month	Day	Duration of suspention time (minutes)
A	2012	April	25	10
B		June	13	51
C		July	10	58
D		August	3	93
E		September	6	66
F	2013	January	27	72
G		February	6	105
H		February	13	14
I		February	20	40
J		March	5	30

2.2 Basic analysis

Figure 1 shows the changes with the lapse of time of number of "tweet". The "tweet" increase from occurrence of railway accident to 60 minutes in each cases. During the railway operation stopped, railway users send information and personal situations around them to "twitter" platform.

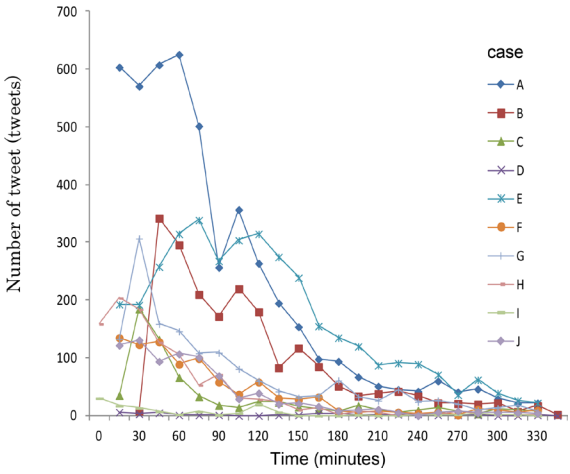

Figure 1 Changes with the lapse of time of number of "tweet"

Figure 2 shows changes with the lapse of time of accumulation of number of "tweet". After the accident occurred, the number of "tweets" is stable in 150 minutes. Average suspension time of railway accident is about 53.9(minutes), but the "tweet" were sending to twitter platform after the railway operation start. This phenomenon means the delay of the railway affects it for a long time.

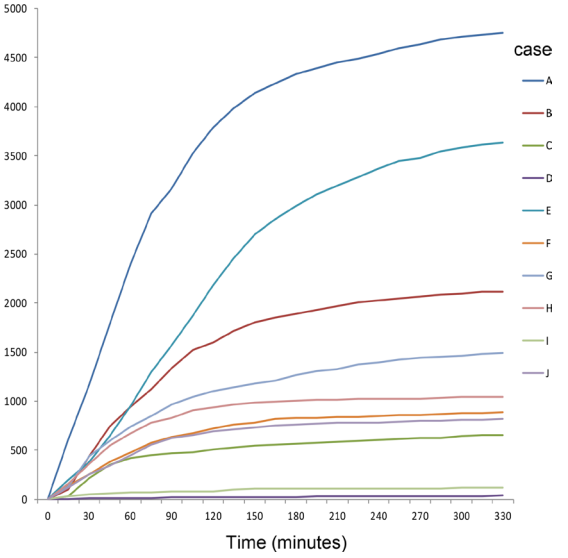

Figure 2 Changes with the lapse of time of accumulation of number of "tweet"

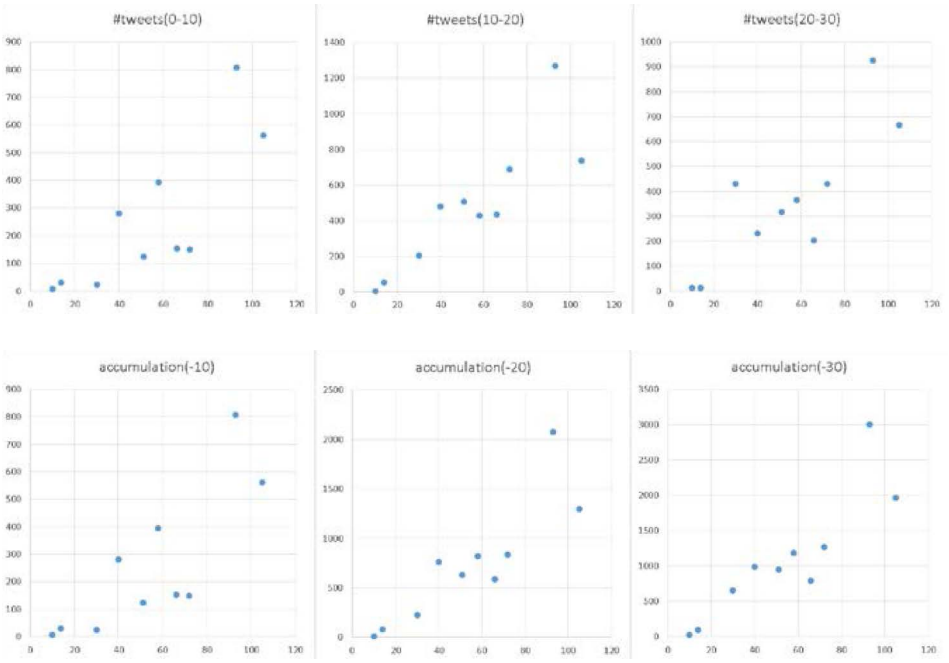

Figure 3 Relationship between number of tweet and suspension time

Figure 3 shows relationship between number of tweet and suspension time. Relationship between number of tweet and suspension time has positive correlation. As a result of this figure, regression analysis can be adopted in this study.

3 Suspension time forecasting model

3.1 Simple linear regression

Figure 4 shows that result of simple linear regression. Tweet data was delimited every five minutes and regression analysis was done. A red dot means observed data, and blue dot means estimated data. As seen in the figure, the reproducibility is high. When the twitter data until ten minutes is used, R is 0.8.

3.2 Exponential regression model

Figure 5 shows that result of exponential regression. Tweet data was delimited every five minutes and regression analysis was done. A red dot means observed data, and blue dot line means exponential regression. As seen in the figure, the reproducibility is high. When the twitter data until 5 minutes is used, R is 0.87 which is better than result of simple linear regression.

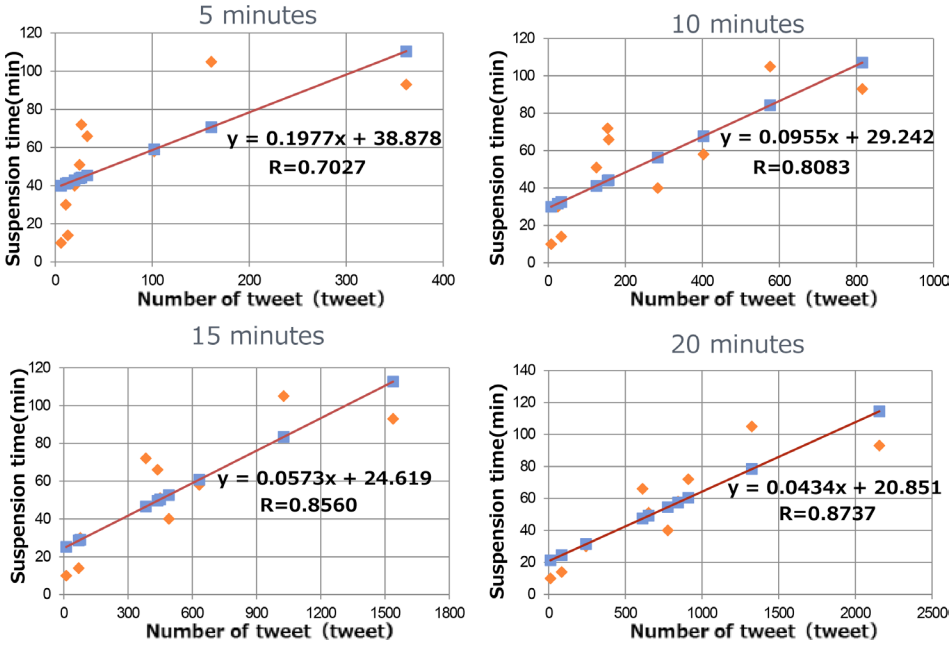

Figure 4 Simple linear regression

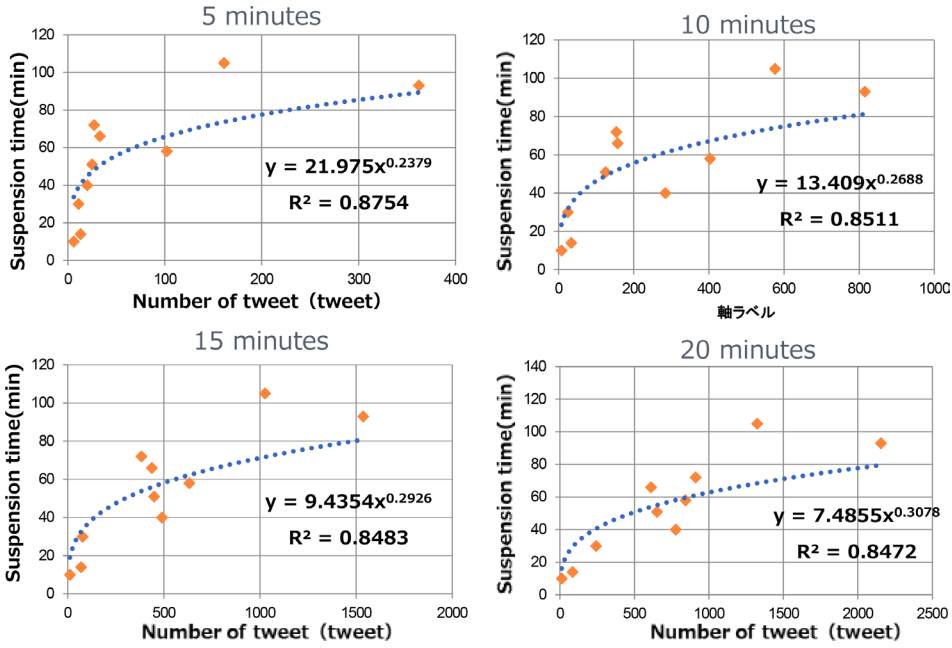

Figure 5 Exponential regression

4 Conclusions

In this study, number and contents of "tweet" are analysed under accidental conditions which are a small railway accident and a railway accident in which a lot of people get influenced. Furthermore, simple linear regression and exponential regression were applied to develop the suspension time forecasting model. Then, it became clear that in case of using simple linear regression, the reproducibility is high. When the twitter data until ten minutes is used, R is 0.8, and in case of using exponential regression, the reproducibility is very high than simple linear regression. When the twitter data until five minutes is used, R is 0.87 which is better than result of simple linear regression. As a result of this study, suspension time of railway accident can be forecasted using twitter data until 5 minutes by exponential regression model.

References

[1] Takada, K., Fujiu, M. & Yokoyama, S.: Perception of the Travel Time Reliability of the Urban Railway Service in Tokyo, Proceedings of the First Conference on Road and Rail Infrastructures – CETRA 2010, 17–18 May 2010, Opatija, Croatia, pp. 873-879.

[2] Takada, K., Fujiu, M.: Study of Willingness to Pay for Reducing Lost Time of Railway Users, Journal of Eastern Asia Society for Transportation Studies, Vol. 8, pp. 1464-1474, 2010.