Although all care was taken to ensure the integrity and quality of the publication and the information herein, no responsibility is assumed by the publisher, the editor and authors for any damages to property or persons as a result of operation or use of this publication or use the information’s, instructions or ideas contained in the material herein.

The papers published in the Proceedings express the opinion of the authors, who also are responsible for their content. Reproduction or transmission of full papers is allowed only with written permission of the Publisher. Short parts may be reproduced only with proper quotation of the source.
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
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, University 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. Vassílios 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
STUDY ON THE AVAILABILITY OF “TWITTER” DATA FOR FORECASTING SUSPENSION TIME OF RAILWAY OPERATION

Makoto Fujiu¹, Kazuyuki Takada²
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>
<thead>
<tr>
<th>Table 1</th>
<th>“Twitter” data (in Japanese)</th>
</tr>
</thead>
<tbody>
<tr>
<td>created_at(JST)</td>
<td>tweet_text_data</td>
</tr>
<tr>
<td>2012/8/3 18:02</td>
<td>強化線3日目終了です！ע downloader 今朝は東上線の遅れに運ばれた都民もいましたが、無事活動することができました。土日は休みなので、ちゃんと休んで氷温強化線に備えましょう！</td>
</tr>
<tr>
<td>2012/8/3 18:04</td>
<td>@nason_kazu_k そうだね？RT 東上線人身事故、運転見合わせ中。朝から渦まったんだ。</td>
</tr>
<tr>
<td>2012/8/3 15:59</td>
<td>「人が死にやすい鉄道」橋、東京上線発。死亡率でトップだそうですね。誰かに話していたら管理、過去のものはあとで説明しますからね。</td>
</tr>
<tr>
<td>2012/8/3 15:58</td>
<td>@dancebook1 東上線は1時間半ほど停まられました(;;)_最近までは事故が多くですね。</td>
</tr>
<tr>
<td>2012/8/3 15:34</td>
<td>@nakadomi123 東上線、人身事故多いね！( Herald &amp; the</td>
</tr>
<tr>
<td>2012/8/3 15:22</td>
<td>お問い合わせ、東上線人身事故など事例からフィードバック</td>
</tr>
<tr>
<td>2012/8/3 15:20</td>
<td>東上線選出情報現在、事故・運転に関する情報はありません。12/08/03 15:18 8駅の運営監視</td>
</tr>
<tr>
<td>2012/8/3 15:18</td>
<td>そうですね、朝、東京東上線にまとまって…</td>
</tr>
<tr>
<td>2012/8/3 15:13</td>
<td>朝の朝鮮新交通局も、amentsの影響で、東上線は通常、渋滞の影響で、東京上線を上げる</td>
</tr>
</tbody>
</table>
| 2012/8/3 15:13 | 今日はバスから説明会に行ったのですが、約2時間前から東上線が運行を止めています。絶望ですね。初めて行く説明会運転は、会社の見解です。

Table 2 | Selected railway accidents |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Case</td>
<td>Year</td>
</tr>
<tr>
<td>A</td>
<td>2012</td>
</tr>
<tr>
<td>B</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>2013</td>
</tr>
<tr>
<td>G</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td></td>
</tr>
<tr>
<td>J</td>
<td></td>
</tr>
</tbody>
</table>
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](image1.png)

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](image2.png)

Figure 2  Changes with the lapse of time of accumulation of number of “tweet”
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
