Traffic-condition Analysis using Publicly-Available Data Sets

Ulrich Dangel¹, Patrick McDonagh², Liam Murphy¹

¹ LERO@UCD, School Of Computer Science and Informatics
University College Dublin, Belfield, Dublin 4, Ireland.
ulrich.dangel@ucdconnect.ie
liam.murphy@ucd.ie

² LERO@DCU, School Of Electronic Engineering
Dublin City University, Glasnevin, Dublin 9, Ireland.
patrick.mcdonagh@dcu.ie

Abstract

In this paper, we introduce some Dublin-specific public traffic data sets and analyse traffic data by linking it with other, non-traffic related datasets. We explain irregularities in observed journey times with weather phenomenas, public events and public holidays. We discuss how the timing of different weather phenomenas influences the observed journey time. By combining different data sources, we can provide reasoning for observed journey times which can be used to explain unexpected traffic patterns, improve capacity planning and aid with other traffic engineering tasks.

Keywords: ITS, Data Analysis, Ireland

1 Introduction & Related Work

Dublin is ranked as the 16th most congested city in Europe which is worse than London and Munich [1]. As a method to control congestion Dublin City Council employs the Sydney Coordinated Traffic System (SCATS) [2] to handle traffic in the city. In 1989, Dublin installed SCATS [3] to monitor and control traffic; the overall monitored junctions increased during the years from 170 in 1997 [4] to more than 700 currently [5]. The overall increase in monitored junctions in Dublin emphasises the importance and necessity of an ITS.

An Intelligent Transportation System (ITS), such as SCATS, provides traffic engineers and system operators with the necessary metrics, such as congestion level [3], to control and monitor the traffic of a road network. The Irish National Transport Authority suggests the use of an ITS for controlling and monitoring large road networks to “enhance operational efficiency and driver information” [6].

The data provided by an ITS helps to monitor and detect problems such as traffic congestion or hotspots. As an ITS is a domain specific monitoring and control system, it does not aide in finding the cause of a problem as of result of events outside of the traffic domain, such as bad weather or large public events. By linking different data sources together, we can find what caused the problem, even if it was caused by events external to the domain.

The intelligent part of an ITS relies heavily on the data and data sources [7]. By using, for example, multiple traffic-related input sources such as induction loops, GPS probes [8] or ultrasonic sensors, a multisource-driven ITS can reduce the deployment costs by combining different sensor types or increase the monitored area by leveraging floating car data for cost-efficient traffic monitoring [7].

Understanding what caused a specific delay or congestion by combining and linking traffic and non-traffic related data sources was discussed in [9] with the focus on the semantics of the data sources. By
developing ontologies, they successfully linked different data sets such as roadworks and maintenance data with public transport data.

In our research, we focus on a general approach for monitoring a distributed system by using different data sets and combine these with other, not directly-related, data sources. Our main research goals are 1) finding irregularities in time series data for a distributed system and 2) helping diagnose irregularities by linking different data sets together. We are using traffic data as an example of a distributed system as such a cyber-physical imposes additional difficulties such as noisy data, regulation, privacy issues as well as communication issues.

Detecting and finding irregularities in time series data offers a broad range of applications such as 1) detecting defective sensors 2) data cleansing 3) verifying simulations and traffic models 4) supporting automated incident detection on traffic data by taking additional information into account.

2 Data Sets

There are a wide variety of public data sets available, from financial data to public health data. In this section, we will discuss some publicly available, local (Irish) data sets, which we will combine later to explain anomalies in traffic data.

Even though not all of the data sets discussed here are directly related to traffic, we will show how the combination and fusion of different data sources can explain traffic patterns and allow for a better understanding of the data and traffic conditions.

2.1 TRIPS Data from Dublin City Council

The Travel-time Reporting and Integrated Performance System (TRIPS) integrates with SCATS and is used to analyse and predict travel times between junctions monitored by SCATS. Dublin City Council publishes the data in real-time to Dublinked [5], a local website focused on publishing local, public sector data.

The published TRIPS data consists of the 1) locations of junctions monitored by SCATS, 2) a collection of defined routes and 3) journey time estimates. A route is a ordered list of links and a link is a ordered pair of positions associated with a junction. The journey time estimates are provided for each individual link of a route and for both directions. The explicit distinction between the travel directions can simplify the data analysis, e.g to compare morning and evening commuting, as shown in Figure 1a. This figure highlights the difference between the two directions, Direction 1 is primarily used in the morning, while Direction 2 is mostly used in the evening. Without this distinction, the spike at around 18:00 for Direction 1 would not be visible.
Table 1: Estimated journey time comparison between TRIPS data and commercial routing services

<table>
<thead>
<tr>
<th>Route</th>
<th>Dublin City Council</th>
<th>Provider A</th>
<th>Provider B</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>12 minutes</td>
<td>11 minutes</td>
<td>13 minutes</td>
</tr>
<tr>
<td>11</td>
<td>5 minutes</td>
<td>6 minutes</td>
<td>5 minutes</td>
</tr>
<tr>
<td>24</td>
<td>14 minutes</td>
<td>15 minutes</td>
<td>13 minutes</td>
</tr>
<tr>
<td>30</td>
<td>14 minutes</td>
<td>13 minutes</td>
<td>12 minutes</td>
</tr>
</tbody>
</table>

2.1.1 Quality

The resolution of the TRIPS data is of a high granularity, as estimates are provided for the complete route as well as for individual links. The data is updated every minute, which allows for the detection of small, fine-grained patterns.

For a small subset of the routes, we verified the estimated travel time with the travel times from proprietary routing services, as shown in Table 1. The estimated times were similar during normal, non-congested times, i.e. in the early morning or evening. To compare and verify the provided estimates during congested time we would have to compare the estimates with real journey times captured by GPS probes [8] or by comparing it to other data, for example Dublin Bus data.

2.1.2 Limitations

Based on this data set, Dublin City Council currently (as of December 2012) monitors at least 1277 junctions with SCATS, though only 717 junctions have an explicit position, while others either have only street names or generic descriptions like ROAD2. This is only a minor issue as each link of the provided routes has a start and an endpoint.

Some of the 45 defined routes in Dublin City also have minor inconsistencies, such as alteration between the sides of the river Liffey, as shown in Figure 1b, or only consist of one link and other minor issues.

2.2 Journey Times from South Dublin County Council

South Dublin County Council (SDCC) publishes journey times on their website for 42 routes. The published journey times consist of a route name, average journey time, journey time and average speed. The route name is a pair of two canonical names; identifying the start and end point without an exact position. Average journey time is the average, historical, estimated time needed to traverse the route while journey time is the current estimated time. It is unclear however, how the journey time and average speed is calculated, e.g. by speed monitoring, induction loops or vehicle tracking with CCTV.

2.2.1 Quality

Depending on the time of day, the journey times are updated with a frequency ranging from five minutes up to multiple hours. Due to this delay, the data has a high variance between each of the data points. The published routes also don’t contain enough information, as the exact route is not provided but only a start and end name, e.g. Heuston Station to Palmerston.

2.2.2 Limitations

SDCC does not publish the data in a machine readable format but on a website targeted for end-users [10], e.g. people travelling alongside one of the specified routes. In order to acquire the provided data, we need to scrape the website and extract the information. The calculations for a route may also be provided at different times, i.e. comparing routes to each others imposes additional challenges due to temporally unaligned data.
2.3 Weather data

The National Oceanic and Atmospheric Administration (NOAA) publishes weather information for all airports in the standardised Meteorological Aviation Routine Weather Report (METAR) format. METAR is a file format for reporting weather information which is widely used in aviation and meteorology.

METAR contains standard weather information like temperature, wind speed and weather phenomena as well as other aviation specific information like ICAO (International Civil Aviation Organization) code, visibility and cloud conditions. As METAR data is used for aviation purposes, the data can be considered to be of high quality.

We decided to use the weather data recorded from Dublin Airport as it is in close proximity to Dublin City Centre. While manually validating extraordinary weather phenomena (snow on January 22nd 2013) in Belfield, Dublin, we noticed a discrepancy in the recorded data even though the linear distance to weather station is approximately 13 kilometers. This can be mitigated by using more weather stations, providing increased local coverage.

2.4 Comparison

We choose to use the provided TRIPS times from Dublin City Council instead of SDCC as the TRIPS data has a higher resolution and is more complete, i.e. it specifies the actual route and not only an imprecise start and end name. The TRIPS data is also updated more regularly than the SDCC journey time data. Even though we are currently focusing on the TRIPS data, we may integrate the SDCC journey time data to get a better overview of the greater Dublin area by combining both data sets.

3 Data Processing

In this section, we discuss an subset of necessary steps to analyse different data sets, how to combine different data sets, as well as different ways to analyse the used TRIPS data.

3.1 Temporal Alignment

Different data sets can have different time resolutions, e.g. the METAR data is provided half-hourly, while the TRIPS data from Dublin City Council is provided minute-by-minute. Other data sets may have no fixed time resolution or specify a time range, e.g. a concert from 19:00 to 21:00.

In order to find corresponding sensor readings from other, temporally unaligned sensors, we need to map the data from one time specification to another. This mapping depends on the sensor type. Observations for different phenomena have a different underlying model; traffic behaves different than temperature, for example. Based on the underlying model, different methods may be employed.

3.1.1 Interpolation

Depending on the nature of underlying data, a simple linear interpolation between the measured data points or more advanced interpolation based on a model may be feasible. For air temperature, we can use a model-based interpolation to reduce the average error rate and improve the overall findings [11].

3.1.2 Time range

While air temperature can easily be interpolated, other data such as weather phenomena is difficult to interpolate. Instead of interpolating the data points themselves, we can transform the measurement to a time range by combining it with the previous and next data point. Equation 1 and 2 shows a simple method to define a new range based on the predecessor and successor of a data point, where \( t_s \) is the new start point, \( t_e \) the new end point and \( t_n \) the time of the \( n \)th measurement

\[
t_s = t_n - \frac{t_n - t_{n-1}}{2}
\]  

(1)
Depending on the observed weather phenomena, the formulas can be adjusted to take additional information or transformations into account and dynamically adjust the period on the next and previous conditions.

3.2 Micro versus Macro Analysis

The TRIPS data from Dublin City Council provides journey time estimates for individual links, as well as for complete routes. This allows us to compare congestion propagation as well as how merging multiple routes onto the same link influences traffic.

Figure 2 shows two links on the same route. Figure 2a (Link 8) shows a link with only one route passing through, while Figure 2b (Link 9) shows a link where multiple routes merge. The journey time on Link 9 may be higher than on Link 8 as after Link 9 three routes are merged traveling towards Dublin City. This means that traffic signals are adjusted to merge these three major routes together and creates congestion, providing a possible reason for the increased journey time.

Even though micro analysis can provide a good insight into specific routes or behaviour, we are currently focused on macro analysis to provide a good overview of pattern throughout a city. With the combination of macro and micro analysis, we can provide a good explanation on specific time pattern caused by hotspots.

4 Patterns

In this section, we will discuss some of the patterns (observed journey times) we detected while verifying and analysing the TRIPS data from Dublin City Council. Currently, this analysis is performed manually, future work will extend this analysis by automating this approach using stochastic methods. We will provide an explanation for these patterns by linking it with external, non-traffic data sources such as weather data and public event data.

4.1 Bank Holiday

Figure 3 shows the journey time for three consecutive Mondays for a specific route. The last Monday in Figure 3 has a significantly lower average estimated journey time than the previous Mondays, as October the 29th 2012 was a bank holiday in Ireland.

The reported journey time for the public holiday matches expectations as the observed route is a commuter route into Dublin city. The reduced commuter traffic, due to closed local offices and businesses, reduce the overall journey time. The slightly increased journey time in the morning could be ascribed to traffic travelling into the city center for shopping and/or to the restrictions imposed by the Dublin Marathon, occurring at the same time.
4.2 Public Event

Bank holidays produce a simple traffic pattern, as the majority of the workforce are not required to work. A slightly less obvious pattern, caused by large public events, is shown in Figure 4. The estimated journey times are for a route in close proximity to the Aviva Stadium, a local sports stadium with a capacity for over 50,000 spectators.

The traffic pattern shown in the first chart is caused by an sold out American Football match in the Aviva Stadium. To cope with the traffic, the police closed some roads at 09:30; at 12:00 the local police closed more roads to cope with the spectators [12]. The temporary reduction of the estimated traffic time could be explained by the partial road restrictions imposed by local police. Directly after the event, at approximately 17:30, there is a large increase in journey time as spectators leave the premises.

The second graph in Figure 4 shows an typical Saturday with no events in the Aviva Stadium. The relatively small and short increase around 13:00 could be ascribed to smaller events in close proximity to the route such as TedX (in the Bord Gais Energy Theatre), or shoppers travelling into the city centre.

The last chart in Figure 4, shows the estimated journey times caused by an sold-out concert at the Aviva Stadium. The gates opened at 17:00 which is visible in the journey time estimates. The decreased journey time after 13:00 can be ascribed to local police authority closing roads around the sport stadium [13]. The morning traffic pattern resembles roughly the previous, normal week but the evening traffic differs a lot due to the sold-out event in the evening.
4.3 Weather

Another pattern we detected is that caused by severe weather. We chose a route from Dublin North to Dublin City Centre as representative of typical commuter traffic. Figure 5 shows journey times on consecutive Mondays; even though on all the Mondays, some weather phenomena was recorded, different journey times can be observed.

For January the 7th (first chart), 14th (second chart) and 28th (fourth chart) rain was registered by the weather station in Dublin Airport. On January 7th, rain was registered throughout the whole day while, for January 14th rain was only registered up until 6 in the morning and later in the afternoon. The increased travel time persists in the first chart over a longer time period than in the second chart. Compared to these two, quite similar charts, rain on January the 28th was first registered at 9 in the morning. The increase, compared to previous Mondays, happened later, around 9 when rain was detected.

On January the 21st snow was registered throughout the whole day. The severe and high increase in journey time can be attributed to snow as it is an uncommon event in Dublin. The increased journey time was not only higher than to the other days but it also lasted for a longer period.

4.4 Findings

We provided causes for specific traffic behaviour and showed that, by using non-traffic related data sources, we can aide and provide reasons for specific irregularities in journey times.

The observed journey times had different characteristics; (1) on the public holiday there was a strong decrease in the estimated journey time throughout the whole day due to reduced overall traffic levels, (2) how large public events influenced the traffic for routes close to the premises but how mitigation employed by the local police helped to reduce the observed travel time and (3) how different weather phenomena, as well as their timing and duration, have a large influence on the journey time.

5 Conclusion and Future Work

We introduced and assessed different, publicly available, local data sets for traffic information and other non-traffic related data sets. We verified the traffic data by explaining and cross-referencing observed journey time patterns with other data sets. We have shown the influence of bank holidays, large public events and weather phenomena such as rain and snow on journey time.

Our current, manual, approach for detecting abnormal behaviour is driven by visualising journey time estimates. This manual process has been proven to be valuable to verify and understand unknown data sets. In order to scale and process a large amount of data, we plan to automate the process of finding irregularities in time series data by creating a model for the data.
In order to find an appropriate model, we plan to use different statistical analysis methods to detect, for example, periodic, shifted data. The analysis of time series data can also help to classify unknown behaviour. We plan to use traffic related data sets as an use case for our approach as the highly distributed nature introduces additional challenges for the data quality and the amount of incurred data.

Our aim is to find a general approach for analysing time series data and cross referencing it with data sets from other domains. As distributed systems grow larger and become more complex, there is a need for automatic detection of irregularities and to fuse performance data with other data sets. We are currently using traffic data as a city resembles a distributed system in the physical world. By cross-referencing data sets, we showed that our approach can provide a better understanding of the observed system and support resource and capacity planning, as well as root cause analysis.

Acknowledgments

This work was supported, in part, by Science Foundation Ireland grant 10/CE/I1855 to Lero - the Irish Software Engineering Research Centre (www.lero.ie)

References