Geospatial Visual Analytics of Traffic and Weather Data for Better Winter Road Management

Yuzuru Tanaka, Jonas Sjöbergh, Pavel Moiseets, Micke Kuwahara, Hajime Imura, and Tetsuya Yoshida

Abstract Sapporo is a city with two million citizens that gets 6 m of snow per year. This means that winter road management is very important for sustaining economic and social activities during the winter. We believe that an exploratory and iterative analysis and visualization approach is useful to support the decision making, to improve the winter road management strategies. We propose using a huge library of tools and services, and a framework that allows users to freely federate tools and services improvisationally ("mash-up") to create custom visualization and analysis environments and to apply these on appropriately selected data sets. Unlike conventional macro analysis approaches, we focus on micro analysis of winter road conditions. We use probe car data, speed readings etc., automatically collected from taxis and private cars. Geospatial visualization of the average speeds of all the road segments shows how different roads are affected by heavy snowfall, by snow plowing, and by snow removal. Combining geospatial visualization with knowledge discovery algorithms is a potential approach in this area. An example would be clustering the road segments based on similarity of the impact snowfall has to group roads into groups that can be maintained using similar strategies.

Keywords Exploratory visual analytics • Geospatial data visualization • Data mash-up • Improvisational federation • Winter road management

Y. Tanaka (✉) • J. Sjöbergh • P. Moiseets • M. Kuwahara • H. Imura • T. Yoshida
Hokkaido University, Sapporo, Hokkaido Prefecture 060-0810, Japan
e-mail: tanaka@meme.hokudai.ac.jp; js@meme.hokudai.ac.jp; moiseets@meme.hokudai.ac.jp; mkuwahara@meme.hokudai.ac.jp; hajime@meme.hokudai.ac.jp; yoshida@meme.hokudai.ac.jp

1 Introduction

1.1 Winter Road Management in Sapporo

Snow plowing and snow removal are fundamental public services for sustaining economic and social activities during the winter in Sapporo. Sapporo is the fifth largest city in Japan and has a population of 1.9 million people, and an annual cumulative snowfall of about 6 m with the maximum depth of snow cover reaching about 1 m. The local government's annual budget for snow plowing and removal is about 15 billion yen. Sapporo is one of few metropolitan cities with this severe snowfall. Others cities with heavy snowfall include Saint Petersburg, Harbin, Montreal, Ottawa, Helsinki, Calgary, Toronto, Syracuse, Anchorage, Buffalo, Rochester, Denver, etc., though none of them come close to Sapporo considering both population and snowfall.

Car traffic is seriously influenced not only by heavy snowfall but also by huge piles of snow on both sides of the roads. These piles (or *windrows*) of snow are created by the snow plowing. In Sapporo, both daily activities of residents and business and industrial activities depend heavily on automobile mobility.

Inadequate winter road management significantly slows traffic flow because of icy roads and roads being narrowed by windrows, which leads to further traffic congestion. Congested traffic will compress the snow and turn it into ice, making the roads extremely slippery; common around pedestrian crossings where cars have to break and accelerate. Repeated melting and freezing of roads by heavily congested traffic will also make the icy roads bumpy. This makes it difficult to steer, and also causes damage to the underside of the cars when it scrapes against the ice.

Studded tires were widely used in Japan, but the use was regulated in early 1990s to solve air pollution problems caused by the dust stirred up by studded tires. As a result, air quality improved significantly. However, it also resulted in several traffic issues in winter; e.g. an increase in car accidents on icy roads, slower traffic flow, increased use of anti-freezing agents and abrasives, and a significant increase of road management costs. Asano et al. estimated the direct and indirect economic losses from the ban on studded tires (Asano et al. 2002). The estimated annual losses were more than 18.5 billion yen in the Sapporo area. The main causes were the increase in driving times and costs, road accidents, and costs for maintenance and management.

1.2 Quantitative Studies of Winter Traffic

Takahashi et al. pointed out that in order to study traffic issues in winter it is important to understand the traffic patterns in a quantitative manner (Takahashi et al. 2004). The traffic patterns during weekdays notably differ from those during weekends or holidays. The traffic in winter also seems to vary greatly depending on specific weather conditions, e.g. snow cover, snowfall, and declining temperatures.

Takahashi et al. used taxi probe car data (taxi GPS data) as an advanced survey method to analyze traffic. They used multiple linear regression (MLR) analysis to explain the reduction in average travel speed (ATS) in winter compared to the ATS in summer in terms of explanatory variables including snow depth, snowfall amount, snowfall amount on the previous day, average temperature, and sunshine duration. Their MLR analysis showed that decrease in average temperature has the largest effect on reduced average travel speed in winter. Their counterintuitive result is that the amount of snowfall during the day or the previous day has little effect on the ATS. This conclusion however depends on their macro analysis, using only the average speed of the whole city.

Munehiro et al. from the same laboratory estimated the time loss along a specific route due to traffic winter congestion, using taxi probe car data (Munehiro et al. 2012). They also estimated benefits of snow removal on the same route. The analysis is still macroscopic.

The number of taxis and private cars from which probe car data can be obtained in real time has increased a lot in Japan. Probe car data are becoming fundamental in quantitative micro analysis of traffic changes caused by snowfall and icy roads.

Probe car data has been used for various purposes. One example of the use of probe car data for disaster management occurred after the Tohoku earthquake and tsunami in Japan in 2011. Car manufacturers in Japan have car navigation systems in the cars as standard, and these can use the driver's cell phone or other means of wireless communication to report back the car location and speed. These data are commonly used to provide services such as information on where there are traffic jams. Many roads were destroyed or blocked by the tsunami and earthquake. Three large car manufacturers (Honda, Toyota, and Nissan) provided the government with their probe car data. This could then be used to determine which roads were still being used, thus not destroyed or blocked, and which were not used, thus needing cleanup or restoration.[1] This information was also useful for planning where supplies could be sent, and other routing.

For other cities with severe snow conditions during winter, there have been scientific and technological research studies using ICT and GIS approaches to improve winter road management since the mid 1990s (Perrier et al. 2006a,b, 2007a,b). The objectives have been to provide heuristic-optimization solutions for many complex planning decisions. The main strategic and operational problems include defining a service level policy, locating depots, designing sectors, routing service vehicles, configuring the vehicle fleet, and scheduling the use of the vehicles. These activities are interrelated, and the effect that each decision has on the other decisions impacts the ability to provide the desired level of service. Since, unlike in Sapporo, probe car data from taxi cars or private cars (including both retrospective and real time data) are not yet available on a large scale, the actual application of the proposed methods in the real world has not been quantitatively well evaluated by e.g. calculating the change in traffic for different decisions.

---

[1]http://www.its-jp.org/saigai/ (in Japanese), accessed 2012-04-10.

Several measures to compare the road maintenance (snow removal etc.) performance was presented in Thill and Sun (2009) and used for comparing performance at two different highway segments in Buffalo, NY, during and after snow storms. These measure use Automatic Vehicle Identification (AVI) technology to collect vehicle speeds, and then use e.g. the relative decrease in speed or the time until the average speed has gone back to normal as performance measures. The paper also gives a good overview of previous research on the impact of snow on traffic conditions, and of research on performance measurements for snow and ice control.

1.3   Our Approach to the Problem

In this paper we use retrospective probe car data from taxis and private cars, and combine them with other data sources such as meteorological sensor data, snow plowing and removal records, complaints to call centers, and social media data from Sapporo. We propose a geospatial visual analytics environment for micro analysis of the relationships between different data sources by integrating geospatial visualization with data mining and clustering algorithms. Our goal is to achieve more efficient winter road management, better service or lower costs, with the help of information technology.

Unlike the conclusion of the macroscopic analysis by Takahashi et al. (2004), our microscopic analysis and visualization of the average speed reduction in each road segment indicates that it heavily depends on the amount of snowfall, the snow plowing, and the following snow removal. Heavy snowfall is immediately followed by snow plowing, which may result in huge piles of snow on both sides of the roads, making them narrow, which in turn worsens the traffic flow. Snow removal may take more than 1 day, but gradually improves and finally restores the traffic flow.

We provide geospatial visualization of the change of the average speed in each road segment, providing a visual environment for micro analysis of how roads are influenced by snowfall, snow plowing, etc. We also apply clustering algorithms to the change of the average speed in each road segment during snowy days (or weeks) to find groupings of road segments based on similarity of how the traffic is influenced by heavy snow. Such a clustering results may give insights into which roads can be managed in similar ways, and other strategic planning decisions.

Black ice areas around intersections are very slippery and dangerous, both for cars and for crossing pedestrians. Some automobile companies include the activation record of the ABS (Anti Lock Brake System) in their probe car data, which can then be used to detect slippery intersection areas. The probe car data we have do not include the ABS activation records. However, a comparison of the speed distribution histograms between winter and summer may tell us which intersections are slippery. Slippery intersections can be expected to have a large peak in speeds from 0 to 5 km/h, while non-slippery intersections would have histograms similar to the histograms for summer.

We believe that the problem of optimized or better snow plowing and removal is not a simple system that can be modeled by a single monolithic mathematical model. It is a complex system composed of many mutually interrelated systems. Therefore, any macroscopic analysis that applies statistical analyses or knowledge discovery algorithms such as data mining to the whole system may miss significant insights. Statistical analyses and knowledge discovery algorithms assume application to well-formed problems that can be mathematically modeled. One reason why we need micro analysis is that the problem at hand is not well-formed nor easy to model.

We believe that we first need to iteratively refine our analyses depending on observations and evaluations of previous analysis results. Trial-and-error exploration of the data using visualization and analysis tools can help us get a better understanding of the problem and what data to analyze, which questions to ask, etc. Once we have gotten this far, the problem can be considered to consist of subproblems, each of which is approximately well formed, and normal statistical analysis tools or knowledge discovery algorithms can then be applied to them.

We propose an integrated geospatial visualization and analysis environment for such exploratory visual analytics. Such a system requires a large library of visualization and analysis tools. It should support improvisational composition ("mash-up") of whatever visualization or analysis environments may be required for a specific problem, i.e. it should allow free recombination of tools etc. Our integrated environment uses the Webble World as its enabling technology. Webble World is a Web-top system version of the Meme Media architecture which was first proposed in 1993 and has been extensively studied since then. Webble World uses visual components called Webbles that can be improvisationally federated together by users to compose complex applications without the need for coding.

In order to bring GIS, statistical analysis tools, knowledge discovery tools, and SNS (Social Networking Services) like Twitter into the Webble World and to make them interoperable, we created a wrapper for ArcView, a generic wrapper for tools written in R and Octave, and a generic wrapper for Web services. This wrapping allows us to improvisationally federate for instance the Twitter service with geospatial visualization of probe car data within no more than a few minutes. The exploratory visual analytics with improvisational federation of visualization and analysis tools in a large library will provide a new integrated environment for micro analysis of how traffic is influenced by snowfall, snow plowing, and snow removal.

2   Enabling Technology

2.1   Webble World

We create our system using a technology called Webbles. Webbles are software objects intended to make sharing and re-use of functionality and services as easy as copying and pasting texts and images is today. Webbles are the latest incarnation of
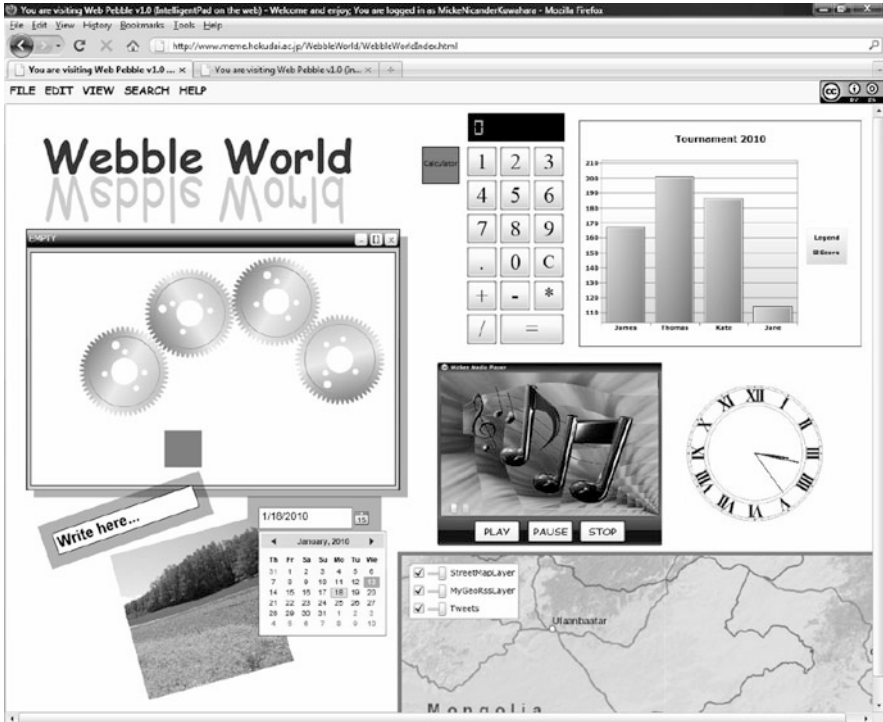
Fig. 1 A few Webbles loaded into a browser in order to display their variety and versatile usefulness. Each single piece or object is a standalone Webble entity; a building block that can be used to interact and collaborate with other Webbles

the IntelligentPad idea. The original idea is based on Richard Dawkins's concept of the Meme (Dawkins 1976), the idea that thoughts, knowledge, and ideas reproduce and mutate in ways similar to what happens to biological genes, and that recombination of ideas can lead to new and perhaps even more useful ideas.

Based on this concept, the Meme Media and IntelligentPad concepts were born (Tanaka 2003). An IntelligentPad is a digital object that can exist in various forms and have multiple purposes, but always has a standardized interaction interface. Several generations of IntelligentPad frameworks have been developed and experimented with during the last decades, but here we focus on the latest version, the Webble World (Fig. 1).

A Webble (Kuwahara and Tanaka 2010) is a customizable object, and Webbles can be loaded from online Webble repositories. Webbles run inside Web browsers (any browser that supports Microsoft Silverlight). Webbles communicate primarily through "slots". A slot is a wrapper that works as the interface for a property of an object or a method it supports. A simple example could be a text label Webble, that can have a slot for the text to display, another slot for the font to use, another for the background color etc.

Slots are named slots because you can plug Webbles together using slots. When you connect two slots the properties of these are synchronized, so when the slot value changes in one object the other object is automatically updated too. A simple example of this could be connecting the "title" slot of a music playing Webble to the "text" slot of a text label to display the title of the song that is being played. When the music playing Webble starts playing another song, the label text will automatically change. For more advanced objects, more interesting two-way communication can also be useful.

Webbles are structured in parent-child hierarchies, which controls the communication structure. This makes it easier to analyze the information flow in complex applications built using Webbles.

There are many Webbles available for use already. Examples include simple components like text boxes, drop down lists, labels, etc.; more complex Webbles include interactive maps, charts, intelligent windows, movie players, etc. Some Webbles allow increased control of other Webbles or the surrounding browser environment, and others allow access to online databases, or allow connecting to Web services.

Freshly created Webbles are called primitive Webbles. Webbles can also be combined together to form more complex objects. Such objects can be saved as a package and are then called compound Webbles. Saving a set of primitive and compound Webbles that are not all connected together is also possible, and this is called a Webble application. Primitive Webbles, compound Webbles, and Webble applications can then be loaded and used by anyone, and it is possible to combine them with Webbles from other sources, or to pick apart complex Webbles and re-use only some parts for new purposes. The idea is that these pluggable components can be used like a "Meme Lego", with simple building blocks that can be recombined to build many different things (Fig. 2).

Primitive Webbles that need to be developed to introduce some functionality not already available in the Webble world are built by programming in Microsoft Silverlight (a subset of Microsoft C#). There are templates available that build the general Webble interface, so only the parts specific to the new functionality need to be written by the programmer. Primitive Webbles like this can for instance be used to build a Webble interface to wrap already existing software. This has been done for instance for the R language for statistical analysis. So you can run R programs inside Webble world, and connect them to the other available Webble components.

Compound Webbles and Webble applications are intended to be easy to create without the need for programming knowledge. The simple interface with slots that you can connect while running in the browser allows all types of users to plug together existing components to solve new problems.

The Webble World is available for use by anyone online.[2]

---

[2]http://www.meme.hokudai.ac.jp/WebbleWorld/WebbleWorldIndex.html
Username: webble, Password: webble.
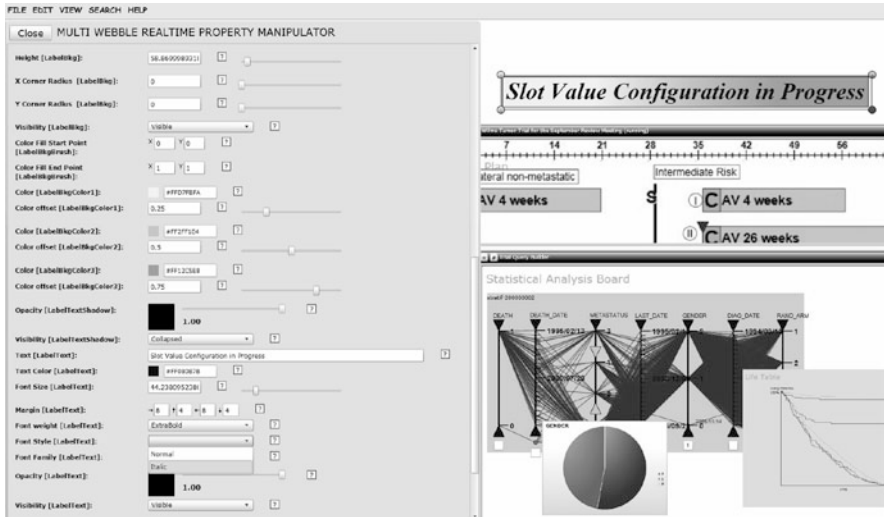
Fig. 2 Just by altering slots values and setting up connection paths between specific Webbles you may build full-fledged rich Internet applications directly online and share them with the world

2.2 Mapping Technology

When dealing with information that is location dependent, it is often useful to display the information on a map, e.g. showing which roads seem to have traffic problems by drawing those roads in red in an overlay over a map of the city. While the same information could be given in other ways, e.g. a table with street names of problem areas, a map is a familiar way to get an overview of the situation quickly.

There are several mapping software toolkits available. One example is the Google Earth (and Google Maps) API[3] from Google. Using a standardized XML format called KML, it is possible to specify for example points, lines, and polygons to be displayed in 3D on a satellite image of the earth. You can also freely position the viewpoint to specify the part of the earth to display, from what angle, the zoom, etc. It is possible to link images, videos, and other data to points on the map too.

Another mapping toolkit is the ArcGIS suite from the Environmental Systems Research Institute's (ESRI). We have wrapped the ArcGIS API for Silverlight[4] into our Webble World framework (Sect. 2.1), and it is thus possible to combine

---

[3]Google. (2012) Google Earth (Version 6.2.1.6014 (beta))
http://www.google.com/earth/download/ge/agree.html (Accessed 6 April 2012).

[4]ESRI. (2012) ArcGIS API for Silverlight (Version 2.4)
http://help.arcgis.com/en/webapi/silverlight/

a mapping tool Webble based on ArcGIS with all the other available software components (Webbles). The Webble wrapper for the ArcGIS has slots for specifying various mapping layers, size of the map, what sections of layers to display, etc.

3   Clustering Taxi Probe Car Data

3.1   Background

Dataset

The taxi probe car dataset was kindly provided to us by Fujitsu Co. LTD. The dataset is based on the information provided by about 2,000 taxi cars, running in the urban area of Sapporo. Roads are split into road segments, with a new road segment starting at each intersection. Statistics are provided for each road segment every 5 min. These include: the average speed, the top speed, the number of cars that passed, the length of the segment, etc. There are around 120,000 road segments in the provided data. We have data for two periods: the *snowfall period*, from January 1st, 2011 to February 7, 2011; and the *non-snowfall period*, from September 19, 2010 to September 25, 2010.

    We investigated the influence of snowfall on the traffic in the dataset. Even when the average speed of a road segment is high in the non-snowfall period, the average speed during the snowfall period might be much lower when the road is covered with snow and ice, or narrowed by piles of snow on the sides. Since the dataset characterizes each road segment in terms of its statistics, road segments can be divided into several clusters based on similarities of these statistics. These groups might then give insights on e.g. roads that are suitable for similar snow removal strategies.

Preprocessing

As detailed above, each road segment is represented by the statistics for the segment gathered every 5 min. In order to get a representation for the segment for 1 day, we simply concatenated the average speed statistics (though other statistics could also be used) of each 5 min period into one high-dimensional vector. This vector does not contain information on the physical properties of the segment, such as the width of the road, the number of lanes, or which other segments are adjacent; it only contains the average speed data. When clustering on the average speed, the a vector would contain 288 speed readings, on for each 5 min period during the day.

    In addition, since the probe car data is transmitted via a radio system from the moving taxis in the real world, the dataset suffers from lots of missing values. In order to calculate the similarities among vectors, it is necessary to fill in the missing

values. How these missing values are assigned affects the results of analysis. As a preliminary experiment, we filled in the missing values of any road segment with the average value of that segment.

Clustering Method

Since each road segment is represented by a high-dimensional vector, we utilized the spherical k-means algorithm (skmeans) (Dhillon and Modha 2001) to clustering the road segments. This algorithm is an extension of the standard k-means algorithm (Hartigan and Wong 1979), and was proposed for the clustering of high-dimensional data such as text. In the standard k-means algorithm, each vector is assigned to the "nearest" cluster in terms of Euclidian distance. However, when the number of dimensions get large, the performance of k-means degrades due to the "curse of dimensionality". In order to remedy the performance degradation, cosine similarity, which is invariant to the number of dimensions, is utilized as a similarity measure in the skmeans algorithm. Thus, each vector is assigned to the "nearest" cluster in terms of cosine similarity.

The true number of clusters is unknown in the dataset. Thus, we varied the number of clusters from 4 to 10 in the following experiments. The results are shown when clustering the data into 6 clusters.

The clustering is only done on speed data. There is no following clustering using the spatial (geographical) information, and the vectors do not contain any such information. This means that the clusters are not (necessarily) spatially connected, and that they can overlap in any way in the physical world.

3.2 Results

The Difference Between Snowfall and Non-snowfall Periods

In order to see if there is any difference between the snowfall and the non-snowfall periods, we created a vector in 288 dimensions for each road segment by concatenating the average speeds of the segment during 1 day (60 min × 24 h / 5 min = 288). A set of such vectors for the road segments is constructed for each day, and clustering with skmeans was conducted for each set of vectors.

Figure 3 shows the clustering results of a day in the non-snowfall period (the number of clusters was set to 6). Normally, the clusters are given different colors and can be overlaid, but here we show the results in grayscale, and show one cluster per image frame. (Figure 7 shows clusters overlaid on top of each other, with one cluster shown in bold for contrast.)

As an example cluster, the second cluster in the top row in Fig. 3, which has all lines shown in **bold**, correspond to the road segments that were clustered together. The depicted lines correspond fairly well with some of the main (arterial) roads
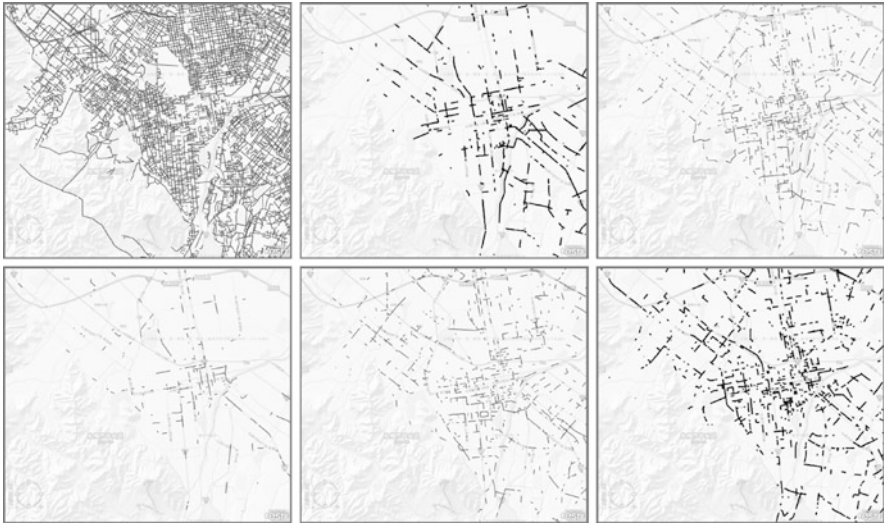
Fig. 3 Clustering result in the non-snowfall period, one cluster per image. Lines in **bold** show clusters with road segments mainly from the main (larger) roads of the city

in Sapporo where many cars run at rather high speed. Note that even though physical connectivity relations among road segments are not utilized in the vector representation, many of the road segments clustered together connect nicely into lines with connected road segments, based solely on their average speed statistics. The last cluster, also shown with bold lines, also contains many road segments from the arterial roads. These two clusters together cover almost all the segments of the arterial roads.

Figure 4 shows the clustering results of a day when heavy snowfall occurred (again, the number of clusters was set to 6). Compared with Fig. 3 we can see that the lines for the clusters with **bold** lines, i.e. the clusters covering the major roads, are more fragmented. Segments from the same road now end up in different clusters more often. This could be caused by traffic jams occurring in some parts of the road segments due to the snowfall, especially in segments where the road has become more narrow than normal because of piles of snow occupying parts of the road after snow plowing. In the next section, we take a closer look at this.

The Difference Before and After the Snow-Removal

The results in Figs. 3 and 4 indicate that the profiles of a road segment are different in snowfall and non-snowfall periods. We believe that this difference in large parts is caused by traffic jams caused by the snowfall. Since piles of snow narrowing parts

Fig. 4 Clustering result for the snow period, one cluster per image. Lines in **bold** show clusters with road segments mainly from the main (larger) roads of the city

of the roads is mostly a problem during rush hour (when few cars are on the road, the road being narrow is not a serious problem), in the following experiment, we focused on data from 7:00 to 10:00 a.m.

In order to highlight the difference between snowfall and non-snowfall periods, we concatenated the average speeds during the rush hour period of 1 day in the non-snowfall period with the speeds during rush hour for 1 day during the snowfall period, thus creating another type of vector of 72 dimensions (60 min × (3 h + 3 h)/5 min = 72). We created such vectors for two days during the snow period: the day directly after a heavy snowfall, and the day two days after the same snowfall.

Figure 5 shows the results for the day just after the heavy snowfall occurred, and Fig. 6 shows the clustering result of the next day (and no snowfall in between). A closer look at an example road is shown in Fig. 7. The **bold** line running vertically is one of the main roads through the city, and right after the heavy snowfall, segments from this road end up in four different clusters. The next day, after snow removal, all the segments in the road are clustered together again.

By comparing these results, we can see that the day after the heavy snowfall had the main roads fragmented, with segments ending up in many different clusters, but the fragmentation is reduced after snow removal. This confirms our belief that snowfall leads to traffic problems and that when the snow is removed, the problems also go away. The results also indicate the importance of microanalysis of traffic to discover the best strategies for road management and maintenance specific to different types of roads.

Fig. 5 Clustering result for rush hour of a day directly after a heavy snowfall. The segments from the major roads end up in several different clusters

Fig. 6 Clustering result for rush hour after snow removal (2 days after the snowfall). The segments from the major roads are generally clustered together
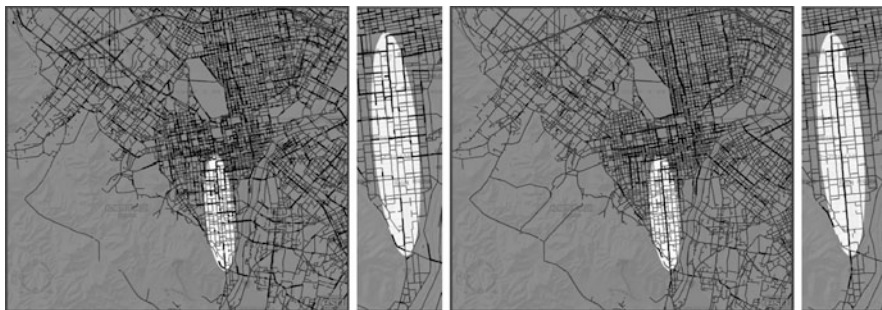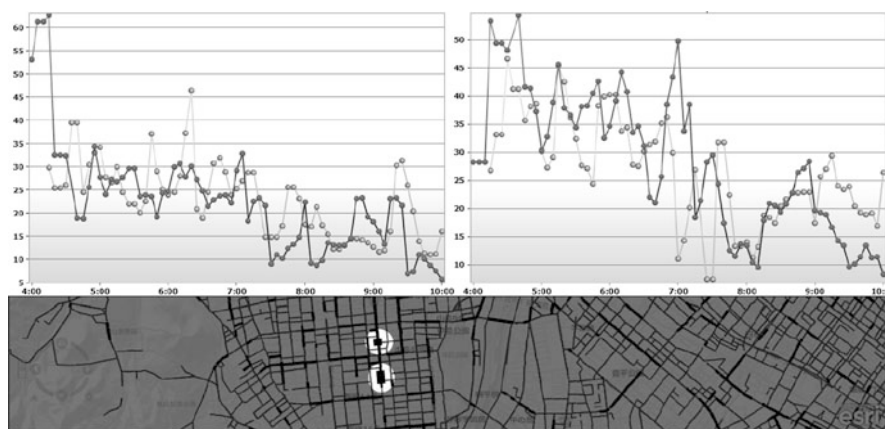
Fig. 7 A closer view of the clustering results of Figs. 5 and 6

Fig. 8 Two road segments that are clustered together during snow free days but end up in different clusters after snowfall are highlighted. Time series of average speeds for the two road segments are shown in two different graphs. The *left half* are readings from the non-snowfall period, and the *right half* are from the snowfall period. Each graph has one *curve* for the day with heavy snowfall (*light gray*) and one for the snow free day after that (*dark gray*)

3.3 Combining the Cluster Results with Other Visualization Methods

In Sect. 3.2, we saw that road segments belonging to the same major road were clustered into the same cluster during the non-snow period and during snow free days of the snowy period. On days with heavy snowfall, and the snowy period overall, segments from the same road ended up in several different clusters, though.

Our system makes it possible to improvisationally federate additional data sources and data visualization tools, to take a closer look at the data or visualize it in other ways. One example of this is shown in Fig. 8. Two road segments that

were clustered together during the snow free day in Fig. 7 (and belong to the same road) are highlighted and drawn thicker than the other road segments. These two segments were clustered into different clusters on the day after the heavy snowfall (Fig. 7).

It is possible to bring up graphs showing the average speed of the road segments as a function of the time of day. The speed during different times of the day is what the clustering is based on. There is one graph for each road segment, and each graph shows two curves. The dark gray curve shows the speeds during the snow free day, and the light gray curve shows the speed during the day with heavy snowfall. We can see that the curves for the snow free day are at least fairly similar in the right half of the graph (the left half are the concatenated data from the non-snowfall period, the right half is from the snowy period), while the light gray curves are very different, as can be expected based on the clustering results in the previous section.

4   Histograms of Speeds in Slippery Intersections

Intersections are especially prone to bad road conditions during the winter in Sapporo. Cars have to break and stop when the light is red, and they have to accelerate and often slip, causing the tires to spin and polish the compressed snow and ice into very slippery ice. The intersections also usually become very uneven and bumpy, making it even more difficult to drive safely and without damaging the car.

Not all intersections are affected the same way, and knowing where there are problems could help prioritize where to go to put more sand (to reduce slipperiness) or file down icy bumps. Here we show speed histograms for cars when they are driving through various intersections around our university. The speed readings are taken from private cars equipped with car navigation systems (collecting the speed readings) and cell phones with subscriptions to a service notifying the navigations system of possible traffic jams etc.

In Fig. 9, speeds are shown as speed histograms, one, in black, pointing up, for the speed in the summer (good road conditions) and one, in gray, pointing down, for the speed in the winter (possibly slippery conditions). The leftmost bar in each histogram is the relative amount of cars driving 0–5 km/h, the second bar from the left is for 5–10 km/h, and so on. The histograms are scaled so that the volume of each histogram is the same, so the bars do not show the number of cars driving slowly but the proportion of cars that traveled slowly.

The rightmost intersection of the two top intersections with histograms is famous for being slippery. We can see that the gray histogram has a larger proportion of cars traveling at slow speeds. In contrast, the leftmost intersection in the lower part of the picture rarely becomes slippery, and the gray histogram there does not have a high proportion of cars at slow speeds.

Fig. 9 Histograms showing the speed in intersections around our university. The *gray bars* are the speed in winter and the *black bars* are the speed in summer. The *bar* farthest to the *left* is 0–5 km/h, the next 5–10 km/h, etc

5 Mashups with Probe Car Data and Twitter

For disaster management, combining many different types of data sources can be useful. We do not necessarily know in advance exactly what data will be of use to us, and the data we want the most may not be available so we may have to make do with combining other sources to achieve the same goal. Here we show an example of combining and visualizing data from many different and diverse data sources.
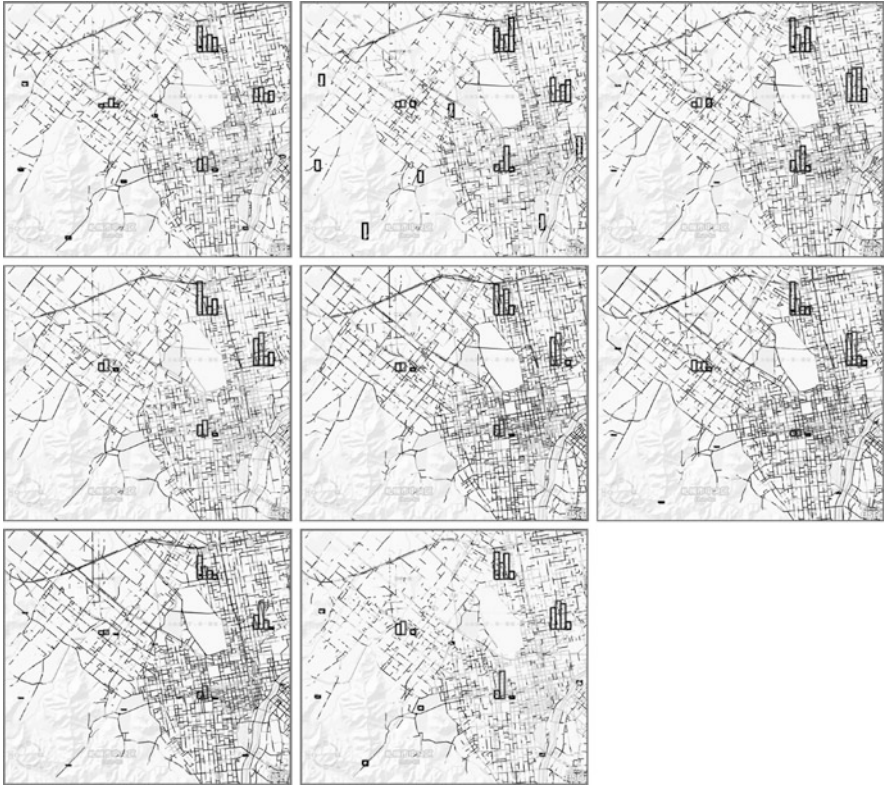
Figure 10 shows a screenshot when visualizing several types of data. Here the visualization has been changed to colors showing up better in grayscale, but in actual use all data sources have clearly distinguishable colors instead. The following types of data are shown:

Probe car data. The streets are colored in white, gray, and black based on probe car data. Where the average speed is roughly the same as the speed in the summer, black is used. On streets where the speed is lower than in the summer gray is used, and when the speed is much worse than in summer the road is colored white. The data is from around 2,000 taxi cars reporting speed readings at each intersection they pass. The day shown in Fig. 10 had fairly bad snow conditions, so the traffic situation was not good and most roads are white or gray.

Weather station data. Snowfall measured at about 50 weather stations located in different parts of the city is shown as black bars standing alone, on top of the location of the weather station. Tall bars indicate large amounts of snow, low bars indicate little snow fall. The weather sensors also collect many other types of data not visualized here.

Fig. 10 A mash-up of many different data sources. The images show the probe car speed data, the most common words from Twitter tweets in Sapporo, weather data, snow removal data, and snow removal complaints to call centers. This visualization was done using Google Earth, not using our Webble World framework

Call center complaints. The amount of complaints from citizens calling the local snow removal call centers or the city call centers are shown as two dark gray bars (a tall bar meaning many complaints) in a group of four bars. The complaints are the two rightmost bars, with the city call centers on the left, and the local snow removal call centers on the right. The data is shown for each ward (district) in the city, and the bars are located in a central location of the ward, usually the location of the ward office.

Snow removal data. Data from the snow removal companies is shown as two gray bars, also in the set of four bars placed in the central ward location. The leftmost bar indicates the road distance plowed by snow plowing vehicles. The second bar from the left indicates the amount of snow removed by trucks taking snow out of the city to specified dumping locations.

Twitter data. The most frequent words mentioned in Twitter tweets that are location tagged as coming from Sapporo are shown in a table. Here there is one global table for the whole city, but displaying local tables for tweets from smaller areas is also possible.

In actual use, showing all the information at once, as in Fig. 10, is probably not very useful since there is too much information being displayed at once. Thus, turning individual data sources on and off has been made easy. This data can then show for instance how the traffic situation was affected by the weather and the snow removal efforts.

Fig. 11 A time series showing the change in the traffic situation after a heavy snowfall. Once the heavy snow hits, the traffic situation becomes bad and there is a large spike in complaints to call centers. The situation then gradually improves, and complaints go down. Then, more snow falls and the traffic situation worsens again. *White lines* indicate streets where the average speed is much worse than in the summer, and *black lines* are streets where the speed roughly the same as in the summer. There is one image per day, *left to right*, *top to bottom*, with the *top left* being January 31, 2011, and *bottom right* February 7

Figure 10 shows the data displayed using Google Earth. The same data can also be shown using the ArcGIS toolkit, which we have wrapped so it can be used in our Webble World framework. The following example scenario is shown using our Webble World.

The example scenario is shown in Fig. 11. The figure shows the days from January 31 to February 7, 2011. On January 31 we can see that some roads are white or gray, so the traffic situation is not as good as in the summer, which is to be expected, but there are not so many complaints coming in. On February 1, there was a very heavy snowfall as can be seen by the many high black bars standing alone all over the city (snowfall measured at weather stations). The traffic situation turns bad, and large parts of the streets in the city are white or gray. There are also spikes in the call center complaints data (the two rightmost bars of the sets of four bars placed together).

The following days there is no or very light snowfall, but the traffic situation does not improve over night. It takes several days before most of the city streets are back to black. The complaints are also still quite high for several days, even though there is no new snowfall. This is of course because there is still a lot of snow left from the snowfall on February 1, as can be seen for instance by the fact that the snow removal companies keep removing very large amounts of snow the following days too (the leftmost bars in the sets of four bars).

The complaints gradually go down as more and more snow is removed, and the traffic situation also goes back more or less to normal. On February 6, the second to last picture, the situation is good. Most of the streets are black (good traffic conditions), and there are few complaints. On February 7 (last day shown), it snows again, and the traffic situation turns bad again.

Since other factors also affect the traffic conditions, adding other data sources can also be useful. Figure 10 also had data from Twitter. This can be useful when something unexpected happens. On the February 7, the Sapporo Snow Festival started. This is a huge event where over a million tourists come to the city, and several big streets around the festival area in the middle of the city are closed off. This of course has a large effect on the traffic flow in the city. When noticing that traffic flow around the area is bad despite snow conditions being good and other streets not having problems, we could bring up a list of common words when people write status updates in Sapporo (or a smaller area). February 7 has "snow" and "festival" among the top five words, and we could guess that this is what is causing the problems.

While the Snow Festival occurs every year and is thus not an unexpected event, a similar approach can be used when unexpected conditions occur. Something that affects a lot of people is likely to be mentioned by many of them. Now that most people are constantly connected through their cell phones, information such as the status tweets on Twitter can be helpful. Our system allows combining data from such diverse sources as cell phone tweets, weather stations, and probe cars.

The strength is that as long as a data source, analysis method, or display method has been wrapped to work in the Webble World environment, connecting them in new ways is very quick and easy. If the Twitter information does not tell us enough, we can plug in a news feed subscription Webble and display words mentioned in news stories about Sapporo instead, with only a few minutes work. Or some other completely different information source.

6   Tweets

In Sect. 5, data collected from citizens writing status updates on Twitter was visualized together with many different types of data. There, we showed the aggregated statistics of all tweets from a specific area, displaying the most frequently mentioned words to see what seems to be going on in that area.

Fig. 12 Location tagged tweets

Instead of showing which words occur in an area, it is also possible to show all the locations of tweets mentioning a specific word. This can tell us where people are when they talk about different things. Not all tweets are location tagged, but many are (actually a small minority of all tweets, but still a fairly large amount in raw number of tweets). Some are tagged only with the general location, e.g. "Sapporo Station", and some have GPS coordinates.

Figure 12 shows two examples of displaying location tagged tweets. There is a text field to input a search string to indicate what tweets you want to visualize, and a map showing the locations of tweets containing this string.

On the left is a picture of the locations of tweets mentioning *"Miku"*. *Miku* is the name of a singing synthesizer persona, and there was a big snow sculpture portraying this pop singer character at the Sapporo Snow Festival. On February 7, this 3 m tall statue fell off its base and injured a tourist. Many tweets mentioning *Miku* can be seen around the West 5 area of the Odori park, which is were the statue was located. There are tweets mentioning the event from other locations too, of course, but there is a clustering effect around the area of the statue.

Similarly, on the right in Fig. 12 is a display of the tweets mentioning the Snow Festival. Most tweets are located in a band from west to east in the Odori park, which is the festival area.

Visualizing where people are mentioning that they are slipping, or that they fell, can tell us where the road conditions are bad. There are however not very many location tagged tweets mentioning such things. Of the 40,000–50,000 tweets tagged as coming from Sapporo each day, only around 800–900 have the specific location (GPS coordinates). Of these, only about one message per day mentions falling or slipping in our data. So slippery roads are not noteworthy or exceptional enough to most people and thus the Twitter feed cannot pick up on it, at least not yet (this might improve when more people start using their GPS enabled smart phones to send the messages). For more severe problems, it would likely be possible to notice them in the messages of ordinary people.

We believe using this type of information source can be very useful. The "end users" of the snow removal services are the citizens, so this information directly

provided by the citizens themselves that can also automatically be collected is hopefully useful, and it has the potential to contain information not covered by the other information sources.

7 Conclusions

Winter road management and maintenance is fundamental in Sapporo for sustaining economic and social activities in winter. Since the influence of snow on traffic is a complex system that is hard to mathematically model as a single system, explorative and iterative analysis and visualization is necessary to support decision making for better management and maintenance strategies.

We proposed using a huge library of visualization and analysis tools and services, and a system framework supporting improvisational federation ("mash-up") of them to compose whatever analysis environment may be deemed necessary based on previous analysis. Unlike conventional macro analysis approaches, we focus on micro analysis of winter traffic conditions using taxi and private probe car data. We showed how visualization of changes in average speed of each road segment shows the influence from heavy snowfall, snow plowing, and snow removal. We also proposed a method to estimate which intersections may have problems with black ice from the probe car data.

We also applied a clustering method to probe car data to classify road segments based on similarity in the effect of heavy snow on the traffic of the segment. This could give insights into management decisions for different segments.

References

Asano M, Hara F, Tanabe S, Yokoyama S (2002) Evaluation of the ten years since studded tires were banned in Hokkaido and future issues, new challenges for winter road service. In: XIth international winter road congress, Sapporo, p 10

Dawkins R (1976) The selfish gene. Oxford University Press, New York

Dhillon JS, Modha DS (2001) Concept decompositions for large sparse text data using clustering. Mach Learn 42(1):143–175

Hartigan JA, Wong MA (1979) Algorithm AS 136: a k-means clustering algorithm. J Appl Stat 28(1):100–108

Kuwahara M, Tanaka Y (2010) Webble world – a Web-based knowledge federation framework for programmable and customizable Meme Media objects. In: The IET international conference on Frontier computing 2010, Taichung, pp 372–377

Munehiro K, Takahashi N, Watanabe M, Asano M (2012) Winter road traffic evaluation using taxi probe data in Sapporo, Japan. In: Proceedings of the 91st TRB annual meeting, Washington

Perrier N, Langevin A, Campbell JF (2006a) A survey of models and algorithms for winter road maintenance. Part I: System design for spreading and plowing. Comput Oper Res 33:209–238

Perrier N, Langevin A, Campbell JF (2006b) A survey of models and algorithms for winter road maintenance. Part II: System design for snow disposal. Comput Oper Res 33:239–262

Perrier N, Langevin A, Campbell JF (2007a) A survey of models and algorithms for winter road maintenance. Part III: Vehicle routing and depot location for spreading. Comput Oper Res 34(1):211–257

Perrier N, Langevin A, Campbell JF (2007b) A survey of models and algorithms for winter road maintenance. Part IV: Vehicle routing and fleet sizing for plowing and snow disposal. Comput Oper Res 34(1):258–294

Takahashi N, Miyamoto S, Asano M (2004) Using taxi GPS to gather high-quality traffic data for winter road management evaluation in Sapporo, Japan. In: Sixth international symposium on snow removal and ice control technology, Spokane, pp 455–469

Tanaka Y (2003) Meme Media and Meme Market architecture. IEEE Press, Piscataway

Thill JC, Sun H (2009) Comparing approaches to winter highway maintenance operations through user mobility performance. J Transp Res Forum 48(1):49–64