Use of Data Mining Technology to Investigate Vehicle Speed in Winter Weather: a Case Study

by

Gang Liu
Graduate Research Assistant
Department of Civil and Environmental Engineering
University of Alberta, Edmonton, Alberta, Canada
Email: gliu2@ualberta.ca

Ling Shi
Graduate Research Assistant
Department of Civil and Environmental Engineering
University of Alberta, Edmonton, Alberta, Canada
Email: lshi5@ualberta.ca

Cheng Lan
Graduate Research Assistant
Department of Civil and Environmental Engineering
University of Alberta, Edmonton, Alberta, Canada
Email: clan@ualberta.ca

Tony Z. Qiu*
Assistant Professor
Department of Civil and Environmental Engineering
University of Alberta, 3-005 NREF
Edmonton, Alberta, Canada T6G 2W2
Tel: 1-780-492 1906, Fax: 1-780-492 0249
Email: zhijunqiu@ualberta.ca

Jie Fang
Assistant Professor
Department of Civil Engineering
Fuzhou University, Fujian, China 350116
Email: jfang1@ualberta.ca

TRB Paper No: 15-2382

Word count: 5403 words + 12 (8 figures + 4 table) * 250 words = 8403 words

*Corresponding author
ABSTRACT

Each winter, Alberta Transportation (AT) spends significant capital on winter road maintenance. It is not enough to use only “time to bare pavement” as the performance measure for winter maintenance. In addition, the databases from various Intelligent Transportation System services in AT are growing explosively every year. New techniques and tools are needed to intelligently transform the abundant of stored data into useful information and knowledge. Previous studies have shown that vehicle speed is a good measure and a fair assessment of winter maintenance operations. This study used the Apriori algorithm, a data mining technique, to investigate the association rules between vehicle speed and various factors. The traffic, weather, and winter maintenance operation data in Alberta are collected and integrated in the SQL server databases. Results of the case study confirmed the expected effects of several weather variables, including snow intensity and pavement surface conditions. The future study will add the collision data into the database for further analysis.
INTRODUCTION

Providing a desired level of service to roadways is the main concern of highway authorities and municipalities in cold regions around the world. Winter road maintenance is essential to eliminate or reduce snow and ice on roadways and to provide safety and mobility during winter seasons. In order to improve the efficiency as well as effectiveness of winter road maintenance and minimize its environmental impacts, there have been many important developments in terms of equipment, materials, technologies, and operations, such as Automatic Vehicle Location (AVL), Road Weather Information Systems (RWIS) and Maintenance Decision Support Systems (MDSS) (1, 2). Concurrently, various performance measures have been developed to understand and estimate quantitatively the benefits of winter maintenance services. According to previous studies, the major direct benefits include fewer crashes, mobility improvement and fuel savings; and the major indirect benefits include increased travel comfort and convenience, reduced insurance claims, and increased productivity (3, 4). One of the most important types of indicators is mobility-related performance measures, such as travel-time savings, speed reduction and recovery and capacity improvement. These parameters are not only easy to monitor from current detection technologies, but also linked to the ultimate outcome of a winter maintenance program.

Since 1996, the maintenance of Alberta’s highways has been carried out by private contractors. Hundreds of people work together to keep Alberta’s 32,000 kilometres of highways clear and open to traffic during the winter (5). Alberta standards ensure that snowplows are on a highway when there are three or more centimeters of snow, or when it becomes icy. Contractors are expected to provide snow and ice control before traffic is heavy. Alberta Transportation (AT) has implemented a number of Intelligent Transportation System (ITS) services throughout the province that have been operational for a number of years, such as the Road Weather Information Systems (RWIS), Automatic Vehicle Location Systems (AVLS), Weight-In-Motion (WIM) sensors, and Automated Traffic Recorders (ATR). These services are able to provide archived and/or real-time data records of information on traffic states, weather conditions, and winter maintenance operations. Alberta deploys significant capital for winter maintenance and for ITS installation and maintenance, but few efforts have been addressed to investigate the benefit of winter maintenance in Alberta (6). The databases of various data sources are growing explosively every year, so new techniques and tools are needed to intelligently transform the abundant of stored data into useful information and knowledge.

It is already known that traffic flow parameters have a close relationship with winter weather and road conditions. In Alberta, it becomes increasingly important to develop the performance measurement system that shows clear linkage between the inputs of winter road maintenance and its outcomes, such as safety and mobility benefits. The objective of this study is to use data mining technology to investigate and identify the mobility benefit of winter maintenance in Alberta from various data sources; particularly, the speed is used as the performance indicator. The study is interested in answering two questions: (a) is there obvious speed reduction and recovery during the winter season? (b) is there an obvious linkage between speed variations, weather conditions, road conditions, and maintenance operations? The remaining parts of this paper are organized as follows. Section two discusses background literature on speed variation during winter seasons. Section three describes the data sources, storage, and study sites. The data mining algorithm used is presented in the fourth section. Section five discusses the analysis results. Concluding remarks are given in section six.
LITERATURE REVIEW

In the past few years, intensive studies have been conducted to investigate the relationship between travel speed and varying weather conditions, as well as road conditions during winter season. The Highways Capacity Manual 2010 states that light snow and heavy snow reduce the free flow speed by 8-10% and 30-40%, respectively, compared with clear and dry conditions (7). Similar conclusions were drawn from some other studies. Lee et al. investigated a total of 954 winter maintenance logs in 24 counties over three winter seasons using regression tree method (8). They found a 20% speed reduction compared to normal winter driving speeds. Agarwal, Maze, and Souleyrette found snow cause 4.17% to 13.46% reduction in average operating speed (9). Robert et al. found reductions in free flow speed for light snow were in the range of 5% to 16%, and 5% to 19% for heavy snow (10). Previous studies had shown that snow storms had a statistically significant effect on speed. Fu et al. stated that the effects varied depending on the type of site and day of the week after investigating 5 sites of urban highway and 17 sites of rural highway (11). Another interesting finding of their study was that the lagged effects were offset by the time and intensity of maintenance operations during and after the event.

Visibility was another factor that had an effect on the traffic speed. It caused 6.62% to 11.78% reduction in average operating speed in study (9) and 10% reduction in freeway flow speed in study (10). Knapp and Smithson found that visibility below 0.4 km reduced average off-peak winter weather vehicle speeds by approximately 6.3 km/h, and snow cover on the roadway lanes reduced average off-peak winter weather vehicle speeds by approximately 11.7 km/h (12). Kwon, Fu, and Jiang also found visibility and road surface conditions had a significant effect on free-flow speed through investigating 25 days of 5-min interval data at two highway locations; however, they stated that snow intensity was found to be significant only when the visibility factor was excluded, which suggested a refutation of these two factors on free-flow speed (13).

In terms of other factors, study (9) found temperature and wind cause 1.21% to 2.10%, and 0.54% to 1.12% average operating speed reduction. Camacho, García, and Belda found the wind speed over 8 m/s affected traffic speed (14). Similarly, they found speed reduction due to snow conditions varied from 9.0 to 13.7 km/h. Study (15) stated drivers reduced their speeds on average of 1.1 km/h for every kilometre per hour that the wind speed exceeded 40 km/h, and the mean speed reduction for all vehicles is 19.2 km/h during the snow events. Kyte et al. found wind speeds higher than 48 km/h caused a reduction in speed by 9.0 km/h, but visibility between 0.2 and 37.0 km caused a very marginal decrease in speeds (16). Similarly, they found snow covered pavement reduced speeds by 10-16 km/h. Cao et al. investigated four highway sites using regression analysis, artificial neural network, and time-series analysis (17). They found that precipitation and road surface conditions have a relatively higher effect on the average traffic speed than other factors, such as temperature and wind speed.

The literature search revealed a good variety of recent studies that have used empirical data to assess the impact of winter weather on traffic speed. Firstly, all past studies have confirmed that winter weather conditions have a negative effect on traffic speed, and the contributing factors, including road surface conditions, visibility, snow intensity, temperature, wind speed, impact of traffic volume, truck percentage, time of day, and others. Due to differing circumstances of study sites (e.g. climate, traffic, and geometry conditions) and data used (e.g. quality and data aggregation interval), the findings were inconsistent in terms of type of significant factors and the size of the effects. Secondly, all the studies need to collect both weather and traffic data over a very long period of time to develop representative models (14). Because of the data heterogeneity and abnormality, the data preparation and debugging process
are usually very important. For example, in study (14), the one-hour traffic data were split into four 15-minute intervals, as the weather stations collected data in 15-min periods. Thirdly, usually linear or non-linear regression models are developed to build the relationship between speed and variables, while some studies divide weather phenomena into several groups and a few studies utilized the time-series analysis. However, it is difficult to take full account of all potential contributing factors (e.g., time of day, traffic incidents, and driver-behaviour characteristics) that contribute to speed reduction, and some variables are also interrelated, which makes the models less reliable or biased (17). For example, Kwon et al. identified two common speed recovery patterns following a snow event: (a) speed recovery is only affected by the road conditions: in this situation, the speed recovers to its free flow speed; (b) speed recovery is affected by both the road conditions and the traffic-flow: in this situation, the speed stops at a certain level before it reaches its free flow speed (18).

Speed reduction and recovery have also been applied to measure the performance of winter maintenance operations. Studies of (8) and (19) confirmed that vehicle speed was a good measure of representing driving conditions during winter weather events and winter maintenance performance alike, and speed recovery duration was also investigated in their study. A sophisticated algorithm was used in the Minnesota Department of Transportation (MnDOT) to identify times of significant changes in speed in order to characterize and measure the return to normal operation (18). Sharifi et al. proposed an algorithm based on the reduction of speed and change in confidence score within vehicle probe data to calculate winter weather road restoration performance measures (20).

Regardless of abovementioned limitations, previous studies have provided insightful and conclusive findings in speed reduction and recovery phenomena during winter seasons and winter maintenance operations, no matter what methods are used to measure the processed data. There have been many methods for mining different kinds of knowledge, such as predictive and descriptive models. The data may reveal important information in different ways if using different types of mining technologies. Association analysis is the process of discovering association rules, which express an association between items or sets of items (21). The aim of the algorithm is to reveal relationships among variables that occur synchronously in a given set of data. In this study, association analysis is conducted to reveal the linkage between various variables during the winter season.

DATA DESCRIPTION AND INTERGRATION
Description of Dataset
The first dataset is the weather and road condition data from RWIS. AT has installed over 75 RWIS stations along major Alberta highways to provide real-time RWIS sensor information, pavement, and local area forecasts to Highway Maintenance Contractors (22). RWIS stations contain special atmospheric sensors on their structures and pavement sensors embedded in and below the road that collect detailed data on weather conditions at and near the road surface. Typical observations made at each station include air temperature, atmospheric pressure, chemical saturation, precipitation status, relative humidity, pavement surface status, pavement temperature, subsurface deep temperature, subsurface shallow temperature, wind average-direction, wind average-speed, wind maximum-speed. All the observations are based on twenty-minute intervals. The pavement surface status reported by RWIS sensors consists of the following index:

- Dry: absence of moisture on the surface sensor.
• Trace moisture: thin or spotty film of moisture above freezing. Surface moisture occurred without precipitation being detected.
• Wet: continuous film of moisture on the pavement sensor with a surface temperature above freezing.
• Chemically Wet: continuous film of water and ice mixture at or below freezing with enough chemical to keep the mixture from freezing.
• Ice Warning: continuous film of ice and water mixture at or below freezing with insufficient chemical to keep the mixture from freezing again.
• Ice Watch: thin or spotty film of moisture at or below freezing.
• Frost: moisture on pavement at or below freezing with a pavement temperature at or below the dew point temperature.

The second dataset is the weather data from Environment Canada (23). It provides an online access to the historical hourly and daily archived weather data at various weather stations across Canada. Typical observations made at each station include air temperature, dew point temperature, relative humidity, precipitation type, visibility, and wind speed, all on an hourly basis with the exception of the precipitation intensity and snow on ground, which are available in daily totals. The reported hourly weather phenomena in winter include snow, snow grains, clear, cloudy, ice crystals, ice pellets, ice pellet showers, snow showers, snow pellets, fog, ice fog, blowing snow, freezing fog, and other.

The third dataset is the traffic data from WIM sensors. AT has installed 20 piezoelectric WIM sensors in six different highway locations across Alberta in 2004 (24). Four out of six locations have four lanes with WIM sensors (divided highways), and the remaining two locations have two lanes with WIM sensors (undivided highways) coming to a total of 20 sensors. Four main categories of parameters are measured using a WIM sensor: weight, speed, vehicle classification, and identification. Observations recorded by a WIM sensor include sequential vehicle record number, site number, lane number, date, time, speed, gross vehicle weight, length, vehicle class, weight, and length from the first axle to the tenth axle. All the mentioned parameters are calculated and assigned to each passing vehicle.

The fourth dataset is from the AVLS for plow trucks, which allows automatic reporting of material use, road conditions, and actions taken by the plow operator during a maintenance operation. AT staff are able to track, in real-time, operations of nearly 650 individual snowplows across the provincial highway network from an internet-based software interface, as well as review activities based upon historic archived data files (5). The monitored parameters include unit name, date, GPS time, location, speed, plowing, sand/salt spreading conditions, billable operating time, and others.

Data Storage and Integration
As shown in Figure 1, a multi-layered architecture is utilized to store and integrate the data from multiple sources: (a) Data Persistence Layer, (b) Data Mining and Analysis Layer, and (c) Data Presentation and Interaction Layer. Microsoft SQL Server 2012, a relational database management system, is used to store and retrieve data. The data from various sources was imported into SQL server. SQL Server 2012 also provides several standard data mining techniques, such as clustering, regression, decision tree, and association rules, which can be utilized to analyze underlying data sets and dig out valuable information from them. The presentation layer employs GIS techniques to present analysis results. This map-based visualization allows end users to gain intuitive knowledge easily, no matter novice, intermediate
and expert users. This presentation layer also provides end users with interactive functionalities, enhancing their understanding and perception of very dynamic phenomena.

![Framework of Data Storage and Integration](image)

**FIGURE 1** The Framework of Data Storage and Integration

---

**Study Sites and Period**

After examining all available data, one study site is selected by considering the data characteristics: (a) detailed traffic data by lane are available; (b) RWIS stations are located proximately at the exact location; (c) one weather station from Environment Canada is located within 10 km. The study site is a four-lane divided freeway facility with a 110 km/hour posted speed limit. It is located on Alberta Highway 2, just south of the City of Leduc, Alberta, Canada (as shown in Figure 2). Highway 2 is among the top five in Canada, in terms of traffic volume and serves as the major economic corridor in western Canada.

Year 2009 is selected as the study period, as all the weather data, RWIS data, and WIM data are valid for almost the entire study period. There are 8,857,899 valid vehicle observations in total in WIM at this location. Each vehicle’s record is correlated to the RWIS data and weather data, according to the time relations.

![Selected Locations](image)

**FIGURE 2** Selected Locations (Red: RWIS Location; Green: WIM Location)
METHODOLOGY

Data mining is the process of using sophisticated data search algorithms and statistical approaches to discover useful hidden information in large datasets in database systems. One of the central themes of data mining is the mining association rules between itemsets in very large databases, which was first presented by Agrawal et al. (21). The basic terms used in this study are as follows:

\[
D = \{T_1, T_2, \ldots, T_n\} \quad \text{A database of transactions}
\]

\[
T_i = \{i_1, i_2, \ldots, i_m\} \quad \text{A transaction contains a set of items}
\]

\[
i_m \quad \text{An item which has an attribute value}
\]

\[
I = \{i_1, i_2, \ldots, i_p\} \quad \text{Items set, } T_i \subseteq I
\]

\[
X \quad \text{An itemset which is a set of items. A transaction } T_i \text{ contains an itemset } X \text{ if and only if } X \subseteq T_i
\]

\[
k\text{-itemset} \quad \text{An itemset with } k \text{ items from } I
\]

\[
\text{Support (} X \text{)} \quad \text{The number of transactions in } D \text{ containing } X
\]

\[
\text{Association Rules} \quad \text{Extracted from frequent itemsets and show correlations among contained items.}
\]

\[
\text{Confidence (} i_A \Rightarrow i_B \text{)} \quad \text{A property of an association rule}
\]

\[
\text{Confidence (} i_A \Rightarrow i_B \text{)} = \frac{\text{Support (} \{i_A, i_B\}\)}{\text{Support (} \{i_A\}\)}
\]

Confidence denotes the strength of implication and support indicates the frequencies of the occurring patterns in the rule. It is often desirable to pay attention to only those rules, which may have reasonably large support. Such rules with high confidence and strong support are referred to as strong rules. The task of mining association rules is essentially to discover strong association rules in large databases.

When several attributes are present, the association rules generation process becomes difficult to deal with because of the large number of possible conditions for the consequent of each rule. To generate the rules efficiently, special algorithms have been developed. One such algorithm is the Apriori algorithm, which is designed to operate on a database containing transactions (21, 25, 26). This algorithm is carried out in this study to learn association rules between vehicle speed and various factors. Apriori algorithm is interested in finding all such rules having high enough support and confidence. The algorithm is shown in Figure 3. It finds frequent itemsets according to a user-defined minimum support. In the first pass of the algorithm, it constructs the candidate 1-itemsets. The algorithm then generates the frequent 1-itemsets by pruning some candidate 1-itemsets if their support values are lower than the minimum support. After the algorithm finds all the frequent 1-itemsets, it joins the frequent 1-itemsets with each other to construct the candidate 2-itemsets and prune some infrequent itemsets from the candidate 2-itemsets to create the frequent 2-itemsets. This process is repeated until no more candidate itemsets can be created. Finally, the frequent itemsets are converted to rules with enough confidence by splitting the items into two, as items in the antecedent and items in the consequent.
CASE STUDY

Snow Event-based Analysis

The daily and hourly weather phenomena information in 2009 is reviewed from a nearby weather station of Environment Canada. Three snow events from December to February are selected to investigate the variation of traffic speed during snow events, as shown in Table 1. Each selected snow event has typical before-snow, during-snow, and after-snow time horizons. Ice crystals appeared during all three events, and fog appeared in event two and event three.

<table>
<thead>
<tr>
<th>TABLE 1 Summary of Weather Phenomena during Selected Snow Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weather Phenomena-Event 1</td>
</tr>
<tr>
<td>Day 1 Clear and cloudy</td>
</tr>
<tr>
<td>Day 2 Cloudy, then snow</td>
</tr>
<tr>
<td>Day 3 Snow</td>
</tr>
<tr>
<td>Day 4 Snow, later ice crystals</td>
</tr>
<tr>
<td>Day 5 Ice crystals, then cloudy</td>
</tr>
<tr>
<td>Day 6 Cloudy, then snow</td>
</tr>
<tr>
<td>Day 7 Snow</td>
</tr>
<tr>
<td>Day 8 Clear</td>
</tr>
<tr>
<td>Day 9 Clear and cloudy</td>
</tr>
<tr>
<td>Day 10 Clear, then freezing fog</td>
</tr>
</tbody>
</table>
Average vehicle speed and pavement surface conditions are calculated or collected for each hour of three snow events. Figure 4 shows the plotted hourly average speed and surface conditions between the beginning and the end of each event.

a) In the snow event in December, the first snow comes at 7:00 AM on Day 2 and the second snow comes at 10:00 AM on Day 6. Before the first snow comes, the pavement surface condition is dry, but the speed starts to decrease. When the first snow comes, the average speed is reduced by almost 30 km/h within 10 hours. Then the speed starts to increase and achieves the normal speed when the first snow ends. The surface condition is ice warning or ice watch during this period. During the weather of ice crystals, there is no obvious speed reduction. When the second snow comes, the speed starts to decrease again.

b) In the snow event in January, the first snow comes at 2:00 AM on Day 2 and the second snow starts at 1:00 AM on Day 4. When the first snow comes, the average speed is reduced sharply by almost 20 km/h within 5 hours. Then the speed starts to increase and achieves the normal speed when the first snow ends. The surface condition is ice warning or ice watch during this period. When the second snow comes, the average speed decreased gradually by about 25 km/h within 35 hours. The surface condition is also ice warning or ice watch during this period. Then the speed starts to increase and achieves to the normal speed when the second snow ends. But the southbound average speed recovers much slower and achieves the normal average speed 20 hours after the second snow ends. On Day 10, when the freezing fog comes, the average speed also starts to decrease.

c) In the snow event in February, the snow comes at 1:00 AM on Day 2. When the fog comes, the speed decreases immediately. The speed decreases continually when the snow comes and is reduced gradually by about 15 km/h within 25 hours. The surface condition is ice warning or ice watch during this period. When the snow ends, the average speed is a little far from the normal speed and achieves the normal speed about 30 hours after the snow ends. The pavement surface almost stays at dry condition.

In summary, it is visually obvious that the vehicle speed starts to decrease when snow comes and starts to increase during the snow event. When the snow ends, the vehicle speed recovers almost to the normal speed. Concurrently, the pavement surface conditions during the snowfall are usually ice warning and ice watch, while it achieves dry status after the snow. It seems the speed reduction and recovery in a snow event has clear a link to the pavement surface condition. However, there are also some speed reduction cases when the snow does not appear. One interesting finding is that speed reduction usually starts before the snow comes. There may be other factors that contribute to this speed reduction; for example, visibility is reduced before snow comes. In addition, the magnitude and steep degree of speed reduction is different between three snow events, which may be caused by the snow intensity. The weather phenomena during winter snow events are usually complicated; therefore, the speed reduction may be caused by multiple factors. Another interesting finding is that southbound and northbound directions of the study site have significantly different speed variations during the three snow events.
FIGURE 4 Speed dynamic before, during, and after snow events.

Note: Solid line: speed; Dashed line: surface condition; Blue line: northbound; Purple line: southbound.

Pavement surface: 3-dry, 4-trace moisture, 5-wet, 6-chemically wet, 7-ice warning, 8-ice watch, 13-frost.
The speed recovery time highly depends on the maintenance activity. Alberta maintenance standards ensure that snowplows are on a highway when there are three or more centimeters of snow or when it becomes icy. Contractors are expected to provide snow and ice control before traffic is heavy. Figure 5 shows the operation of one snowplow during nine days, which is responsible for serving the study location. Four operation statuses are investigated: plow down and up, wing down and up, prewet of and off, billing on and off. The AVL system at AT requires operators to use a billing switch to indicate when their work was part of a maintenance contract. At the end of a workday, the AVL system creates an automated bill that must be reviewed by the vehicle operator before submission. We can see that the maintenance operation starts once the snow event comes, and will last for several days after the snow event which may depend on the snow intensity.

**FIGURE 5 Maintenance operations before, during, and after Snow Events**

The Box Plots is investigated to identify the traffic flow and capacity trends. The data are selected and categorized into three groups: before, during, and after snow event. Traffic flow data at three snow events are collected, and there are 672 hours data for each group. As shown in Figure 6, the capacity decrease significantly during the snow events, and they can recover to normal states after the snow events.

**FIGURE 6 Traffic flow distributions before, during, and after snow events**
Association Rules Analysis

Data Preparation

In order to conduct the association rules analysis, each factor considered in this study will be categorized into several groups. Therefore, the distribution analysis is conducted to determine the thresholds of each group and the minimum thresholds of support.

Firstly, the speed data are analyzed in several time horizons: summer (from the last snow in May to the first snow in October), winter (from the first snow in October to the last snow in May), the time when pavement surface condition is ice warning, and the time when pavement surface condition is ice watch. Figure 7 shows cumulative frequency and a box plot of the speed distribution. Table 2 shows the descriptive summary of statistics. Although the speed limit is only 110 km/h at this location, the median speed and mean speed are both above the speed limit in summer and winter. Even when the pavement surface is ice watch or ice warning, the median speed and mean speed are also almost equal to or above the speed limit. In addition, it is clear that there is obvious difference between speed distribution in summer and winter season. The median speed tends to be reduced by 2 km/h in the winter season. The standard deviation of speed in winter is larger than in summer. There is significant difference between the speed reductions during ice warning and ice watch conditions. The median speed is reduced about 8 km/h in ice warning conditions, compared with the summer time. Finally, we can also see the directional speed distribution is slightly different.
Note: The red line represents distribution in winter, and the blue line represents distribution in summer.

**FIGURE 7 Speed Distributions of Two Locations**

**Table 2 Descriptive Summary of Statistics of Speed Distribution (km/h)**

<table>
<thead>
<tr>
<th>Index</th>
<th>Explanation</th>
<th>25th Percentile</th>
<th>Median</th>
<th>75th Percentile</th>
<th>Mean</th>
<th>Standard Deviations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Northbound in summer</td>
<td>111</td>
<td>117</td>
<td>123</td>
<td>117</td>
<td>9.7</td>
</tr>
<tr>
<td>2</td>
<td>Northbound in winter</td>
<td>109</td>
<td>115</td>
<td>122</td>
<td>115</td>
<td>11.0</td>
</tr>
<tr>
<td>3</td>
<td>Southbound in summer</td>
<td>109</td>
<td>116</td>
<td>120</td>
<td>113</td>
<td>12.9</td>
</tr>
<tr>
<td>4</td>
<td>Southbound in winter</td>
<td>106</td>
<td>114</td>
<td>119</td>
<td>111</td>
<td>14.1</td>
</tr>
<tr>
<td>5</td>
<td>Northbound in ice warning</td>
<td>99</td>
<td>109</td>
<td>117</td>
<td>107</td>
<td>14.9</td>
</tr>
<tr>
<td>6</td>
<td>Northbound in ice watching</td>
<td>107</td>
<td>114</td>
<td>120</td>
<td>113</td>
<td>11.5</td>
</tr>
<tr>
<td>7</td>
<td>Southbound in ice warning</td>
<td>99</td>
<td>109</td>
<td>116</td>
<td>106</td>
<td>15.3</td>
</tr>
<tr>
<td>8</td>
<td>Southbound in ice watching</td>
<td>107</td>
<td>114</td>
<td>119</td>
<td>111</td>
<td>13.5</td>
</tr>
</tbody>
</table>
The weather data are investigated in two time horizons: whole year and the winter season (from the first snow in October to the last snow in May). Figure 8 gives the summary statistics of the weather-related parameters. It is worthwhile to mention that distributions of humidity, pressure, temperature, pavement temperature, surface condition and total daily snow are significantly different between the whole year and winter season. This makes it intuitive that these parameters may contribute to the speed reduction in winter season.

Considering the distribution of the factors, the single itemsets used in the algorithm are described in Table 3. The minimum t thresholds of support and confidence are set to 0.1% and 70%, respectively.

**TABLE 3 Single-Item Set Description**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Group</th>
<th>Item</th>
<th>Variable</th>
<th>Group</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed (km/h)</td>
<td></td>
<td></td>
<td>Humidity</td>
<td></td>
<td>d</td>
</tr>
<tr>
<td>&gt;110</td>
<td>A</td>
<td></td>
<td>&gt;90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(100, 110]</td>
<td>B</td>
<td>(%)</td>
<td>(60, 90]</td>
<td>e</td>
<td></td>
</tr>
<tr>
<td>(90, 100]</td>
<td>C</td>
<td></td>
<td>(30, 60]</td>
<td>f</td>
<td></td>
</tr>
<tr>
<td>(80, 90]</td>
<td>D</td>
<td></td>
<td>≤30</td>
<td>g</td>
<td></td>
</tr>
<tr>
<td>≤80</td>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pressure (mB)</td>
<td></td>
<td></td>
<td>Air Temperature (°C)</td>
<td></td>
<td>h</td>
</tr>
<tr>
<td>&gt;1000</td>
<td>F</td>
<td></td>
<td>&gt;-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(975-1000]</td>
<td>G</td>
<td></td>
<td>(-20, -1]</td>
<td>i</td>
<td></td>
</tr>
<tr>
<td>≤975</td>
<td>H</td>
<td></td>
<td>≤-20</td>
<td>j</td>
<td></td>
</tr>
<tr>
<td>Pavement Temperature (°C)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;55</td>
<td>I</td>
<td></td>
<td>[0-90]</td>
<td>k</td>
<td></td>
</tr>
<tr>
<td>(-1, 55]</td>
<td>J</td>
<td>(degree)</td>
<td>(90-180]</td>
<td>l</td>
<td></td>
</tr>
<tr>
<td>(-10, -1]</td>
<td>K</td>
<td></td>
<td>(180-270]</td>
<td>m</td>
<td></td>
</tr>
<tr>
<td>≤10</td>
<td>L</td>
<td></td>
<td>(270-360]</td>
<td>n</td>
<td></td>
</tr>
<tr>
<td>Average Wind Speed (km/h)</td>
<td></td>
<td></td>
<td>Maximum Wind Speed (km/h)</td>
<td></td>
<td>o</td>
</tr>
<tr>
<td>&gt;70</td>
<td>M</td>
<td></td>
<td>&gt;70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(40, 70]</td>
<td>N</td>
<td></td>
<td>(40, 70]</td>
<td>p</td>
<td></td>
</tr>
<tr>
<td>≤40</td>
<td>O</td>
<td></td>
<td>≤40</td>
<td>q</td>
<td></td>
</tr>
<tr>
<td>Daily Snow (cm)</td>
<td></td>
<td></td>
<td>Atmospheric Precipitation</td>
<td></td>
<td>r</td>
</tr>
<tr>
<td>&gt;15</td>
<td>P</td>
<td></td>
<td>Dry</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>(3, 15]</td>
<td>Q</td>
<td></td>
<td>Wet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤3</td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Road Surface Condition</td>
<td></td>
<td></td>
<td>Visibility (m)</td>
<td></td>
<td>t</td>
</tr>
<tr>
<td>Dry</td>
<td>S</td>
<td></td>
<td>&gt;500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wet</td>
<td>T</td>
<td></td>
<td>(200, 500]</td>
<td>u</td>
<td></td>
</tr>
<tr>
<td>Trace moisture</td>
<td>U</td>
<td></td>
<td>(100, 200]</td>
<td>v</td>
<td></td>
</tr>
<tr>
<td>Chemically Wet Ice Warning</td>
<td>V</td>
<td></td>
<td>(50, 100]</td>
<td>w</td>
<td></td>
</tr>
<tr>
<td>Ice Watch Frost</td>
<td>W</td>
<td></td>
<td>≤50</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Traffic Volume (veh/h)</td>
<td></td>
<td></td>
<td>Date and Time</td>
<td></td>
<td>y</td>
</tr>
<tr>
<td>&gt;3600</td>
<td>Z</td>
<td></td>
<td>Night hour</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3000, 3600]</td>
<td>a</td>
<td></td>
<td>Weekend</td>
<td>z</td>
<td></td>
</tr>
<tr>
<td>(2000, 3000]</td>
<td>b</td>
<td></td>
<td>Weekday</td>
<td>~</td>
<td></td>
</tr>
<tr>
<td>≤2000</td>
<td>c</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FIGURE 8 Histogram and Cumulative Distribution of Weather-Related Parameters
**Results Analysis**

Table 4 shows the results of association rules, the consequents of which are speed or pavement surface condition. The following statements can be drawn from the association rules:

- If the pavement surface condition is dry, the temperature is above -1°C, the wind speed is smaller than 40km/h, and the snow intensity is smaller than 3cm, the confidence of speed is above 110km/h is about 70%.
- The snow intensity and pavement surface condition have a strong relationship with the speed between 90 km/h and 110 km/h. The confidence is between 60% and 65%.
- The pavement condition of ice warning is associated with the temperature and snow intensity, and the confidence is about 61%.
- The dry pavement condition is associated with the temperature and snow intensity, and the confidence is about 80%.
- Visibility has a very great impact on the travel speed.
- There is no association rule between speed reduction and pressure, humidity and wind direction.
- There is no association rule between speed and night hour, weekend, weekday and traffic volume.

<table>
<thead>
<tr>
<th>TABLE 4 Results of Association Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rules</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>R1</td>
</tr>
<tr>
<td>R3</td>
</tr>
<tr>
<td>R5</td>
</tr>
<tr>
<td>R7</td>
</tr>
<tr>
<td>R9</td>
</tr>
<tr>
<td>R11</td>
</tr>
<tr>
<td>R13</td>
</tr>
</tbody>
</table>

**CONCLUSIONS**

This study investigated the association rules between speed and various factors. The results of this study can be applied in quantifying the mobility benefits of winter road maintenance and explaining the different contributing factors to the speed variations. In the future, the collision data will be added into the database for further mining.

**ACKNOWLEDGEMENT**

The authors would like to express thanks to Alberta Transportation for providing data for this study. The contents of this paper reflect the views of the authors and not necessarily the view of Alberta Transportation.
REFERENCES


Liu, Shi, Lan, Qiu, and Fang


