

When TEDDY meets GrizzLY: Temporal Dependency Discovery for Triggering Road Deicing Operations

Céline Robardet
Université de Lyon, CNRS
INSA-Lyon, LIRIS, UMR5205,
F-69621 Villeurbanne, France

Vasile-Marian Scuturici
Université de Lyon, CNRS
INSA-Lyon, LIRIS, UMR5205,
F-69621 Villeurbanne, France

Marc Plantevit
Université de Lyon, CNRS
Université Lyon 1, LIRIS,
UMR5205, F-69622
Villeurbanne, France

Antoine Fraboulet
HiKoB,
F-69100 Villeurbanne, France

ABSTRACT

Temporal dependencies between multiple sensor data sources link two types of events if the occurrence of one is repeatedly followed by the appearance of the other in a certain time interval. TEDDY algorithm aims at discovering such dependencies, identifying the statically significant time intervals with a χ^2 test. We present how these dependencies can be used within the GrizzLY project to tackle an environmental and technical issue: the deicing of the roads. This project aims to wisely organize the deicing operations of an urban area, based on several sensor network measures of local atmospheric phenomena. A spatial and temporal dependency-based model is built from these data to predict freezing alerts. A demo of our system is available¹.

Categories and Subject Descriptors

H.2.8 [Database Management]: Data Mining

Keywords

Sensor data analysis, temporal dependencies.

1. INTRODUCTION

Every winter, deicing or snow-clearing of roads is an important issue that impacts on the local economy, the public finances and the environment. From one hand, city authorities are encouraged to deploy significant means for road deicing as the road network is a key infrastructure for a highly connected and just-in-time economy. Therefore, roads must

¹http://liris.cnrs.fr/vasile-marian.scuturici/video/kdd_demo.mp4

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

KDD '13, August 11–14, 2013, Chicago, Illinois, USA.

Copyright 2013 ACM 978-1-4503-2174-7/13/08 ...\$15.00.

be kept clear to preserve the local economy. On the other hand, deicing expenses constitute an important part of the public works budget that covers the deployment of substantial means: dedicated road agents and patrols, snowplows and dump trucks spread tonnes of salt on slick roads. For instance, Canada spends \$1 billion on snow removal while USA and France spread respectively almost 16 million and 2 million tonnes of salt every winter. Moreover, road salting may have an important negative environmental impact [2, 4]. Indeed, deicing salts leach into the soils where the ions may accumulate and eventually become toxic to the organisms and plants growing on them. The chemicals can also reach ground water in concentration that is toxic to the ecosystems. The contamination of the ground water may result in health impact to urban population. Deicing operations plays havoc with technical equipment: salt tends to cause corrosion, rusting the steel used in most vehicles as well as the threaded rods used in bridge concrete.

Deicing of roads must therefore be organized wisely in order to limit its negative environmental, technical and health impacts. To this end, road operators rely on weather forecasting. However, the weather alerts are on the scale of an entire urban area whereas topographic (e.g., hill) and urban (e.g. parks, buildings) disparities can cause differences in temperature and freezing phenomena. Consequently, many roads are processed without this being necessary and some slick roads are not deiced because no freezing alert was triggered at the urban scale. To enhance deicing operation management, and process only slick roads, the weather forecasting must be done at a smaller scale. It consists in considering short-lived atmospheric phenomena that are smaller than mesoscale, about 1 km or less [1].

In this paper, we present the GrizzLY (a.k.a. Grand LYon) project and its first results. It aims to implement this strategy of weather forecasting at a small spatial scale for the Lyon urban area (Grand Lyon, France). It consists in the deployment of HiKoB wireless sensors to monitor roads and their use to improve the snow removal and salting of the city. Eight sensors were spread over the urban area and located in the neighborhoods known to be deicing sensitive. Each sensor provides real time information on in-pavement temperature combined with outdoor air temperature and relative

humidity. A new technique of temporal dependency discovery [5] is used on these data and provides a two hours prediction model that makes possible to trigger freezing alerts in a more accurate way than classical weather forecasting does. Notice that this work is the first to consider data mining techniques in micro-scale meteorology where simulation techniques are widely used.

2. GRIZZLY AND TEDDY SYSTEM

An overview of GrizzLY [3] and TEDDY system is provided in Figure 1. The Lyon urban area is instrumented with eight sensor networks that produce streaming data. Three months of data are to date at our disposal. We use them to build and validate a freezing alert prediction model based on temporal dependencies. In the following, we describe the smart road sensor network, the temporal dependency technique used, the freezing alert model construction and the graphical user interface that displays, in real time, the freezing alerts as well as the freezing events.

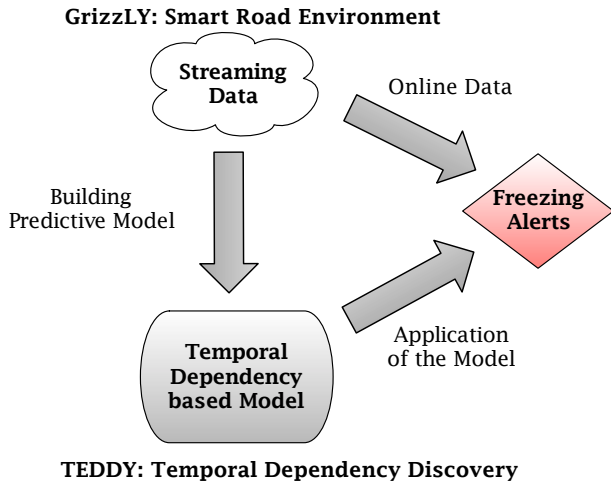


Figure 1: Overview of GrizzLY and TEDDY system.

2.1 Smart Road Sensor Network

HiKoB provides to the Grand Lyon its unique wireless sensor network technology to power its winter service decision making process. The solution is deployed on eight locations spread on the Grand Lyon area. The locations have been selected for their diversity from the environment and weather context point of view: dense urban locations as well as more rural locations at the vicinity of the city.

Each deployment site is equipped with ultra low power wireless sensor nodes that are in charge of collecting outside air temperature and relative humidity similarly to a classical weather station. Four nodes are also cored in the roadway to measure inner road surface temperature as well as temperature at 10 cm and 15 cm below the surface to build a temperature gradient. The road temperature used in this experiment is the average of the inner surface temperature from the four sensing points. These sensors provide measures every 30 seconds.

All measures are directly sent to a HiKoB gateway that connects the local wireless sensor network either to the Grand Lyon metropolitan network or 3G network which then connect to the Internet. Streamed data are sent to a cloud

computing infrastructure using a REST API and can then be made available for computing, visualization or data exploitation.

2.2 Temporal Dependency Discovery

TEDDY (a.k.a TEmporal Dependency DiscoverY) is a data mining tool designed to extract temporal dependencies between multiple sensor data sources [5]. It consists in discovering inter-stream relations that link two types of events if the occurrence of one is repeatedly followed by the appearance of the other in a certain time interval. TEDDY temporal dependencies are robust to the temporal variability of events and identifies the time intervals during which the events are dependent.

TEDDY is based on the following characteristics: (1) the time-point event streams monitored by the sensor nodes are transformed into state streams where events are considered to change the internal state of each sensor node. A node n_k is in state X_k at time t ($state(n_k, t) = X_k$) iff the last event before t of node n_k is X_k . Actually a single geographical site may monitor several phenomena. Therefore, we defined *site states* A as the combination of several node events, $A = X_1 \wedge \dots \wedge X_\ell$. For example, when the air temperature is negative and the route temperature below the frost point value, the site is in *freezing* state. (2) Each stream site state A is defined by the set of time intervals where A is active: $active(A) = \{[a_i, a_{i+1}[\mid a_i \leq t \leq a_{i+1}, state(n_k, t) = X_k, k = 1 \dots \ell\}$ and is characterized by their length $len(A) = \sum_{i=1}^{\#A} (a_{i+1} - a_i)$. (3) The dependency $A \rightarrow B$ of two states A and B is evaluated on the basis of the intersection of their active time intervals: $len(A \cap B) = \sum_{i,j} len([a_i, a_{i+1}[\cap [b_j, b_{j+1}[)$ (4) B can undergo two types of transformation in order to maximizing its intersection with A :

- B can be shifted by α time units, $\alpha \leq 0$:

$$B^{[\alpha, \alpha]} = \{[b_j + \alpha, b_{j+1} + \alpha[\}$$

that is to say each of its time interval is translated in the past of α time units, and we have $len(B) = len(B^{[\alpha, \alpha]})$.

- Or B can be shifted in the past and slightly extended:

$$B^{[\alpha, \beta]} = \{[b_j + \alpha, b_{j+1} + \beta[\}$$

with $\alpha < \beta \leq 0$, and in that case $len(B) < len(B^{[\alpha, \beta]})$.

Extending the intervals makes the temporal dependency measure more robust to the inherent variability of the data. The strength of the dependency is evaluated by the proportion of time where the two states are simultaneously active over the active time period of A :

$$confidence(A \xrightarrow{[\alpha, \beta]} B) = \frac{len(A \cap B^{[\alpha, \beta]})}{len(A)}$$

The information conveyed by this rule is “the occurrence of state A is often followed by the occurrence of state B through a transformation $[\alpha, \beta]$ ”.

A dependence is considered valid if it satisfies a Pearson’s chi-squared test of independence that defines a minimum confidence threshold that depends on $[\alpha, \beta]$. However, a huge number of time shift intervals may exist that result in

valid temporal dependencies and many of them are redundant, depicting the same phenomenon several times. That is why TEDDY focuses in most interesting time shift intervals that lead to a high confidence value and that are as specific as possible with respect to the inclusion relation. To that end, the interval $[\alpha_1, \beta_1]$ is considered to dominate the interval $[\alpha_2, \beta_2]$ if $[\alpha_1, \beta_1] \subseteq [\alpha_2, \beta_2]$ and

$$1 - \frac{\text{conf}(A \xrightarrow{[\alpha_1, \beta_1]} B)}{\text{conf}(A \xrightarrow{[\alpha_2, \beta_2]} B)} < 1 - \frac{\text{len}(B^{[\alpha_1, \beta_1]})}{\text{len}(B^{[\alpha_2, \beta_2]})}$$

that is to say the loss of the confidence measure of $A \xrightarrow{[\alpha_1, \beta_1]} B$ is less than the reduction of its active time period and thus $\text{len}(B^{[\alpha_2, \beta_2]} \setminus [\alpha_1, \beta_1] \cap A)$ is almost 0. Thereby, TEDDY avoids the classical problem of pattern flooding in data mining.

TEDDY extracts all the valid and non-dominated temporal dependencies and their associated time intervals by a level-wise enumeration of the time shift intervals. The confidence value of each interval is computed once at the most. It uses a monotonic property on the confidence measure and a lower bound on the minimum confidence support to early prune the search space. It makes also use of an upper bound on the confidence measure, whose complexity is $O(1)$, to avoid unnecessary computation of the confidence whose complexity is linear in the number of intervals. Finally, it also takes advantage of the transitivity of the dominance relationship in the identification of non-dominated temporal dependencies.

TEDDY algorithm is implemented in standard C++. In [5], we conducted an extensive experimental study of both synthetic and real-world data streams from smart environments equipped with various kinds of sensors (cameras, motion sensors, etc.). These experiments demonstrate that the pruning techniques are very efficient and speed up TEDDY running time by a factor that varies between 2 and 60. A qualitative analysis of the output shows that TEDDY produces a small set of non-redundant dependencies that accurately describe the phenomenon captured by the data.

2.3 Freezing Alert Model Construction

To construct the freezing alert model, we use the data provided by HiKoB sensor sites between November 2012 and January 2013. A first analysis of these data show a strong spatial variability among the temperature measured over the sites. The difference between the air temperature of the sites and the one of the weather station of the city varies between 0.18°C and 1.5°C in average: sites that are close to the weather station have the lower average.

Using these data, we build a model based on the temporal dependencies that link site states. A site state is defined by the five following attributes:

1. The site location name (*ID*),
2. The period of the day (*POD*): 12 a.m.-6 a.m., 6 a.m.-12 p.m., 12 p.m.-6 p.m. or 6 p.m.-12 a.m, one of the four day periods,
3. The road temperature (*RT*): the greatest integer less than or equal to the road temperature,
4. The sign of the road temperature gradient (∇RT): +1 if *RT* is increasing, -1 if *RT* is decreasing, 0 otherwise.

The road temperature gradient at time t (in seconds), $\nabla RT(t)$, is evaluated by $\nabla RT(t) = \frac{RT_{(t-600)} - RT_{(t-0)}}{600}$ with $RT_{t-x} = \text{AVG}(\{RT(t) \mid 600 + x \leq t < x\})$. It evaluates the evolution of the temperature over last two 10 minutes periods,

5. Freezing condition (*FZ*): 1 if the weather conditions are right for it freezes, that is air temperature is less than 0 and *RT* is less than or equal to the frost point value; 0 otherwise.

Site states mainly rely on the road temperature. But as this value is strongly influenced by day / night phenomenon, we also take into account this information as well as its trend. As we want to predict glaze ice on the road, the freezing condition are also considered. Notice that during the considered period, freezing condition occurred during 886 hours among the 14 178 hours of measurement.

TEDDY is used to extract all valid and significant temporal dependencies among the site states. To take into account the road operators logistical and technical constraints, and make the model effective in the deicing decision process, the freezing alert must be triggered at least two hours before a freezing event. Therefore, TEDDY computes temporal dependencies with $-4\text{hours} \leq \alpha \leq \beta \leq -2\text{hours}$. Notice that the construction of the model takes less than one second. We obtain 1567 valid and significant temporal dependencies among which 1320 conclude to a freezing condition site state. For example, we obtain the following rule: *when the site Lacassagne has a road temperature equals to 1°C , with a decreasing gradient and no freezing conditions during the night, then SainteFoy has a road temperature equals to 0°C , with a decreasing gradient and freezing conditions during the morning and this appears between 2 and 4 hours later.*

2.4 GrizzLY and TEDDY Demo

Once the temporal dependency-based model has been constructed, we use it to trigger freezing alerts on new data that can either be fetched using an online RESTfull API, or loaded from a file for demonstration or replay purposes. A graphical user interface makes possible to visualize the current sensor values as well as the triggered alerts (see Figure 2). The sensor locations are displayed on the Grand Lyon city map². The current temperature is represented by a bargraph. When a freezing alert is triggered, the site is surrounded by a square box. A bold blue circle then signals if a road freezing episode is currently happening. By clicking on the sensor location on the map, the end-user can obtain the detail of the temporal dependencies that are triggered in the current freezing alert.

3. PRELIMINARY RESULTS

The aim of our temporal dependency-based model is to predict freezing weather condition. However, these events occur in 6.24% of our data, and half of these events occurred in last third of the data, that is during January 2013. Therefore, to evaluate the efficiency of our temporal dependency-based model, we use the last third of the data to construct the model and we validate it using the 2 first thirds of the data.

Table 1 reports the precision and recall for our model. Our system is set to predict a freezing condition period at least

²We use a map provided by www.openstreetmap.org.

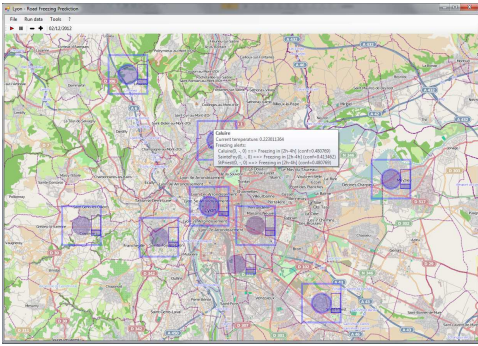


Figure 2: Road freezing prediction in Grand Lyon.

two hours ahead its occurrence. Since the average time interval size of computed temporal dependencies is 17 minutes, a freezing alert is considered valid if an actual freezing appears in the prediction interval or within the four following hours. Indeed, the salt spread on the road remains active at least four hours. Most of the sites have high precision and recall values. Those that have the smallest values are located on the periphery of the urban area. Result variability is explained in the following.

Table 1: Performance of our temporal-dependency based model for triggering freezing alerts.

Site	Precision	Recall
Caluire	0.74	0.67
Craponne	0.35	0.43
Limonest	0.64	0.77
Meyzieu	0.46	0.91
StFoy	0.51	0.9
StPriest	0.77	0.85
Wilson	0.74	0.86

Considering the temporal dependencies that conclude to a freezing state, we have two types of locations: (i) The *predictive locations* that appear in the premise of many temporal dependencies concluding on freezing states. Besides, these locations appears much less in the conclusion of a temporal dependencies; (ii) The *predicted locations* involved in the conclusion of many temporal dependencies depicting freezing phenomena while appearing in the premise part of few such rules. These types of locations can be explained by the prevailing winds. On one hand, the two predictive locations are Craponne and Limonest that are the most westerly sites. During the studied winter, the dominant wind was northwesterly and cold waves often went from West to East. Consequently, this explains why freezing alerts are not very well predicted in Craponne and Limonest locations (no more northwesterly location enabled to trigger a freezing alert that concludes to these locations) while these locations are good cold sentinels. On the other hand, one of the most predicted location is Meyzieu, which is the most eastern one. Other locations are more balanced with regards to their appearance in the left or right hand sides of the temporal dependencies.

Besides providing good performance for triggering freezing episodes, our system produces new actionable insights

that are very hard to obtain using simulation models considering the whole urban area. Furthermore, considering both subareas and short time intervals (prediction every ten minutes) makes possible to highlight *trajectories* of freezing alerts. Such insight is very useful to organize and optimize the deicing operations (e.g. the road operators know that 4 areas must be deiced in a given order).

Lastly, the distribution of the locations in our temporal dependency based model can be used to place new deployment site. For instance, if the road operators want to deploy a new site to enhance the freezing alert triggering on the two most westerly locations, this new deployment site should be placed in the West.

4. CONCLUSION

This paper presents a system that relies on a deployed multi-site sensor network and a data mining technique to predict freezing alerts. We demonstrate that the fruitful combination of sensor network data with a model built with a generic data mining technique makes possible to build predictive models from observed data. Preliminary results on a 3 months period including winter season are promising. These models combines temporal information gathered from measured time series obtained in geographically spread sites over a dense urban area. The presented system is the first attempt, to our knowledge, to apply data mining technique to micro-scale meteorology. The obtained model is a spatio-temporal predictive model that can be used by experts in the fields to make added value from smart environments. Such an approach can then be used in other contexts, in particular in structural health monitoring.

Acknowledgement

This works has been realized in partnership with Grand Lyon and partially funded by the Région Rhône-Alpes and by ANR FOSTER ANR-2010-COSI-012-02.

5. REFERENCES

- [1] P. S. Arya. *Introduction to micrometeorology*, volume 79. Academic press, 2001.
- [2] M. Bester, E. Frind, J. Molson, and D. Rudolph. Numerical investigation of road salt impact on an urban wellfield. *Ground water*, 44(2):165–175, 2005.
- [3] Ludovic Broquereau, HiKoB. Urban traffic management and winter services: wireless sensor networks power smarter decisions. 9th ITS European Congress, June 2013. Dublin, Ireland.
- [4] M. Meriano, N. Eyles, and K. W. Howard. Hydrogeological impacts of road salt. *Journal of contaminant hydrology*, 107(1):66–81, 2009.
- [5] V.-M. Scuturici, M. Plantevit, and C. Robardet. Mining state dependencies between multiple sensor data sources. Technical Report RR-LIRIS-2013-006, LIRIS, <http://liris.cnrs.fr/publis/?id=6030>, 2013.