AUTOMATIC PREDICTION OF ICY CONDITIONS ON ROADS USING A LS-SVM CLASSIFIER

S. Beguin * **G. Bastin** * **V. Wertz** *

* *Université catholique de Louvain - CESAME*
4 av. Georges Lemaître
B-1348 Louvain-la-Neuve - Belgium
Email: {beguin,bastin,wertz}@auto.ucl.ac.be

Abstract: This paper describes a method for predicting the presence or absence of ice on the road. The method is based on a Least Squares Support Vector Machine applied to data from the road in Wallonia (Belgium). It is shown that including a prediction of the air temperature given by a meteorological center in the model helps having better accuracy. In this application, 95% accuracy have been achieved for a 3 hours prediction horizon, and 92% for 6, 12 and 24 hours horizon. *Copyright © 2005 IFAC*

Keywords: Applied Neural Networks, Artificial Intelligence, Decision Making, Machine Learning, Prediction Methods, Classification

1. INTRODUCTION

A few years ago, the roads in Wallonia (southern part of Belgium) have been equipped with 51 meteorological stations that measure meteorological data such as air temperature, subsurface temperature, road state, wind direction etc.

The data are collected and stored in a database managed by the MET (Ministre des Equipements et des Transports). In recent years, some of them were already used to help the road maintenance division in its decision taking process w.r.t. the need of spreading salt on the roads to prevent ice forming. However, recent advances in "intelligent techniques" lead us to study new applications based on these data.

Among possible applications is the construction of an expert system which would act as a computer-aided decision making tool to estimate the need to spread salt on the roads to prevent the apparition of ice. Our work aims at studying the feasibility of such a project.

In the literature, most studies focus on the prediction of the air or surface temperature among others (Hertl and Schaffar (1998), Hertl and Schaffar (1999)) but not on actually predicting ice, or provide a short-term analysis (Shao (1998), Shao and Lister (1996)). In this paper, we used a LS-SVM nonlinear classifier to directly predict the state of the road (icy or not) up to 24 hours ahead.

This paper is organized as follows: we first present the data we worked on. We then show the quality of LS-SVM classifiers when predicting the state of the road. Finally we present the improvement obtained when using a prediction of the air temperature provided by a meteorological center.

2. DATA

Data include 51 sets of 21 variables sampled with a period of 6 minutes, recorded during three winter periods. Among those variables, we find: the air
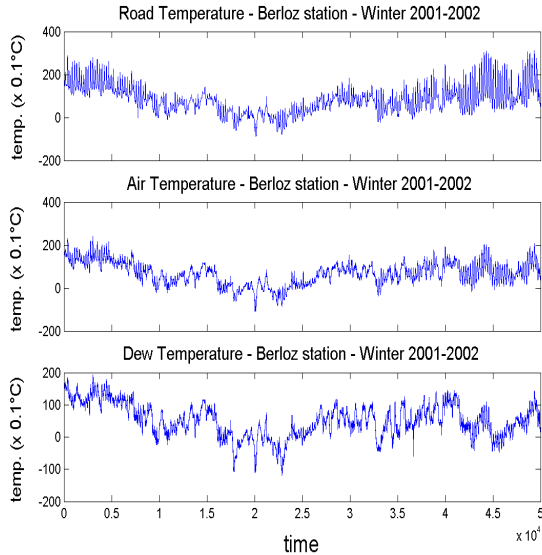
Fig. 1. Some of the variables (Berloz station, 2002)

temperature, the temperature of the road, the hygrometry of the air, etc.

In Figure 1 which presents the road temperature, the air temperature and the dew temperature of the Berloz station, from 1-10-2001 to 30-04-2002, we clearly see a seasonal trend.

2.1 Jalhay Station

One major difficulty in predicting the state of the road is the small number of data when ice is present (it is quite a rare event compared to absence of ice). This could lead us to bad results in classification.

Moreover, data do not explicitly mention whether the absence of ice is natural or due to previous spreading of salt. One solution would be to obtain information about the moments and the places where salt were spread on the roads. This information is unfortunately not available.

The solution we then found was to use data from only one station where ice occurs more often and which is less served by the salting trucks because of its less accessible location. Indeed, one of the stations (Jalhay Station) is located at one of the highest points in Belgium, where icy conditions occur more frequently, and presents much more moments of ice than the others.

This is why we will focus on data from this particular station when building a classifier in order to predict the state of the road (icy or not).

3. LS-SVM CLASSIFIER TO PREDICT THE STATE OF THE ROAD

The problem which we will focus on is the prediction of the state of the road (icy or not), which is one of the variables at our disposal.

A first obvious observation is that there is no ice on the road if the road temperature is strictly positive. We can then perform a pre-classification by computing a prediction of the temperature of the road (for example with a simple linear model). If the prediction of the temperature of the road is above 0^oC (or above a threshold corresponding to the prediction error of the linear model), we can predict that there won't be any ice on the road. If the prediction of the temperature of the road is lower than the threshold, the problem is more complex and we suggest to use a classifier. To build our classifier, we then only used data for which the temperature of the road was below 0^oC.

3.1 LS-SVM Classifier

We use a LS-SVM classifier (Least Squares Support Vector Machines) (Suykens and Vandewalle (1999)). LS-SVM is one particular instance of SVM (Support Vector Machines). In this section we first define SVM classifiers, and we then explain the difference of LS-SVM classifiers with respect to SVM classifiers.

3.1.1. SVM Classifier SVM algorithm tries to obtain a hyperplane to optimally separate different classes. For separable data, the SVM finds the hyperplane with the maximum Euclidean distance to the nearest point. For nonseparable data, the SVM maps the input vectors into a high dimensional feature space, and then constructs an optimal separating hyperplane in this higher dimensional space.

Given a data set of N data pairs $\{y_k, x_k\}_{k=1}^N$ where $x_k \in \Re^n$ is the k-th input pattern and $y_k \in \{-1, 1\}$ is the k-th output pattern, a SVM classifier is of the form:

$$y(x) = sign[\sum_{k=1}^N \alpha_k y_k \Psi(x, x_k) + b] \qquad (1)$$

where $y(x)$ is the prediction of the state of the road for a new input vector x $\in \Re^n$ (1 if ice, -1 if not), x_k are the support vectors (belonging to the learning set, that is the data set used to build the classifier), α_k are support values, and b is a real constant (Suykens and Vandewalle (1999)).

The functions $\Psi(.,.)$ are Gaussian kernels of the form:

$$\Psi(x, x_k) = exp^{-\frac{\|x - x_k\|_2^2}{\sigma^2}} \qquad (2)$$

The SVM associated to such functions is called a gaussian kernels SVM.

We can also express the SVM classifier as:

$$y(x) = sign[w^T \varphi(x) + b] \qquad (3)$$

where $\varphi(.)$ is a non linear function which maps the input space into a higher dimensional space and is such that:

$$\Psi(x_k, x_l) = \varphi(x_k)^T \varphi(x_l) \qquad (4)$$

For separable data, one assumes

$$\begin{cases} w^T \varphi(x_k) + b \geq +1, & \text{if } y_k = +1 \\ w^T \varphi(x_k) + b \leq -1, & \text{if } y_k = -1 \end{cases} \qquad (5)$$

which is equivalent to

$$y_k[w^T \varphi(x_k) + b] \geq 1, k = 1, ..., N \qquad (6)$$

For nonseparable data, a slack variable ξ_k is used:

$$\begin{cases} y_k[w^T \varphi(x_k) + b] \geq 1 - \xi_k \\ \xi_k \geq 0, & k = 1, ..., N \end{cases} \qquad (7)$$

SVM classifiers are obtained as the solution to the following optimization problem:

$$min_{w,b,e} J_{LS}(w, b, e) = \frac{1}{2}w^T w + \gamma \sum_{k=1}^{N} \xi_k \qquad (8)$$

subject to the constraints (7).

3.1.2. LS-SVM Classifier   To build a LS-SVM classifier, the problem is simplified because of the use of equality constraints instead of inequality. A set of linear equations has to be solved rather than a quadratic programming problem; the solution can be obtained in a finite number of steps.

If e_k is defined as:

$$e_k = y_k - (\sum_{k=1}^{N} \alpha_k y_k \Psi(x, x_k) + b), \qquad (9)$$

the LS-SVM optimization problem is written as:

$$min_{w,b,e} J_{LS}(w, b, e) = \frac{1}{2}w^T w + \gamma \frac{1}{2} \sum_{k=1}^{N} e_k^2 \qquad (10)$$

subject to the constraints

$$y_k[w^T \varphi(x_k) + b] = 1 - e_k, \ k = 1, ..., N \qquad (11)$$

The LS-SVM classifier involves two parameters to be optimized, which are σ (the width of the gaussian kernels which cover the input space) and γ (the regularization factor, allowing to avoid obtaining a too local model). The tests were made using the LS-SVMlab Toolbox (Pelckmans et al.).

3.2 Selection of input variables

We select some input variables which have been shown to be highly correlated to the state of the road: past values of the temperature of the road (up to 24 hours ago), the past value of the road temperature 5cm below the surface and of atmospheric radiance along with hygrometry and dew temperature.

3.3 Error Criteria

We consider the two following error criteria:

$$VV = \frac{\sharp \text{ good predictions of ice}}{\sharp \text{ occurrences of ice}} \qquad (12)$$

$$NN = \frac{\sharp \text{ good predictions of 'no ice'}}{\sharp \text{ occurrences of 'no ice'}} \qquad (13)$$

The criterion VV is called the sensitivity, and the criterion NN is called the specificity.

The difficulty in finding a good classifier is the necessity to optimize these two criteria at the same time. It is certainly important to correctly predict the occurrence of ice for reasons of security, but it is also important to correctly predict the occurrence of 'no ice' to avoid an excessive spreading of salt (for ecology and economy reasons).

3.4 ROC curve

One way to express the results of a classifier is the ROC Curve (Receiver Operating Characteristic curve). The ROC Curve is a 2D graph representing along one axis the criterion VV (sensitivity), and along the other axis the criterion 1-NN (1-specificity).

One ROC curve corresponds to some fixed values of the parameters σ and γ. Each point of the curve corresponds to a different value of the threshold b (cf. equation (3)). It gives an idea of the trade-off we can obtain between specificity and sensitivity (in this case: between a good prediction of ice and 'no ice'). Any increase in sensitivity will be accompanied by a decrease in specificity.

The accuracy of the test depends on how well the test separates the group being tested into two classes. Accuracy is measured by the area under the ROC curve. An area of 1 represents a perfect test; an area of 0.5 represents a worthless test.

3.5 Data treatment

If we denote T_e the horizon (in hours), $y(t)$ the state of the road at time t, and $u_i(t)(i = 1...n)$ the input variables at time t, the data set we use is composed as shown in Table 1:

Table 1. Data set

$u_{1...n}(t - T_e)$	$y(t)$
$u_{1...n}(t - T_e + 6min)$	$y(t + 6min)$
$u_{1...n}(t - T_e + 12min)$	$y(t + 12min)$
$u_{1...n}(t - T_e + 18min)$	$y(t + 18min)$

Each line of Table 1 corresponds to one moment (icy or not), and contains information about the value of the input variables T_e hours earlier. As we have one data every six minutes, we then use all the available information.

We used data of the Jalhay station. At the Jalhay station, there were 2810 moments of occurrence of ice and 91042 moments of occurrence of 'no ice' during the winters 2001-2002 and 2002-2003. Using all of these moments to build a classifier takes a lot of computation time and does not allow to validate the classifier on a new data set. We then randomly select a subset of the moments of ice of a fixed size (say N), and randomly select a subset of the moments of 'no ice' of the same size N (in order to avoid difficulties due to big differences in the proportions of the two classes). This forms the *learning set*, the data set which is used to build the classifier. Then, to validate the classifier, we use the $(2810 - N)$ remaining moments of ice, and $(2810 - N)$ randomly selected moments of 'no ice'. This is the *validation set.*

For a fixed number N of moments of ice in the learning set, randomly selecting different subsets of data gives similar results, as shown in Figure 2.
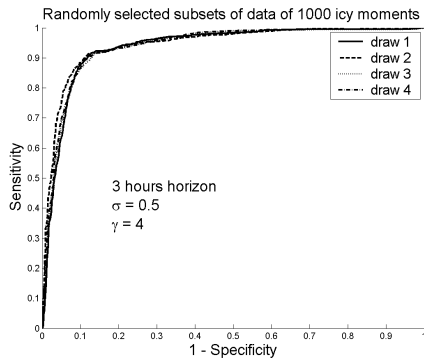

Fig. 2. ROC Curves obtained when using different randomly selected subsets of data

We decide to select $N = 2000$ moments of ice in the learning set, because it represents a good trade off between selecting a large number of points (in order to build a good classifier) and keeping enough points in validation. Figure 3 compares the results obtained when using different numbers of moments of ice in the learning set.
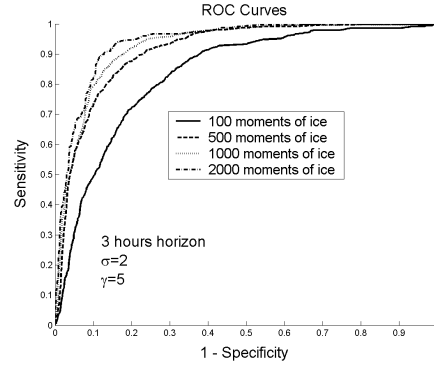

Fig. 3. ROC Curves obtained when using different numbers of moments of ice in the learning set

3.6 Maximizing the area under the ROC curve

As we saw above that the more the area between the ROC Curve is close to 1, the more the classifier is accurate, we maximize the area under the ROC curve in order to select optimal values for the two parameters σ and γ. We first test a large range of values of (σ, γ), going from very small to very large values. We then determine the range of values of (σ, γ) which gives the largest area under the ROC curve. Finally we refine our search in this range of values.

While maximizing the area under the ROC curve according to γ, we can observe that for a small fixed value of σ, the area decreases for increasing values of γ. On the other hand, for a larger fixed value of σ, the area increases for increasing values of γ (cf. Figure 4).
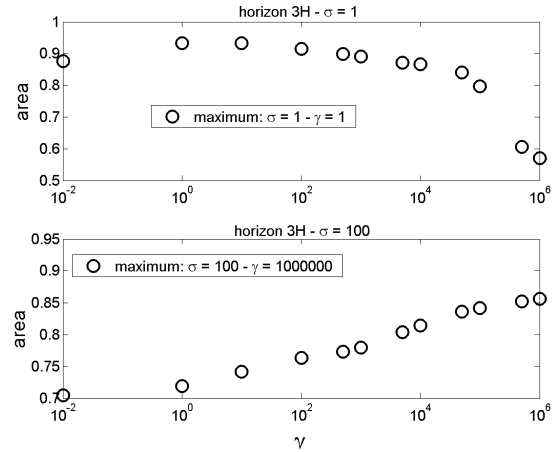

Fig. 4. Evolution of the area under the ROC curve according to gamma for fixed values of sigma

This can be easily interpreted. On one hand, the smaller the σ parameter, the more the model is likely to be too local (the classifier being able to class the learning set perfectly, but giving bad results when classifying new data). On the other hand, the parameter γ is a parameter of regularization: a large value of γ will force the model to be well-fitted to the learning set, a too large value of γ also leading to a too local model.

All this means that for a small value of σ (which represents a risk of having a too local model), we can't afford a too large value of γ, which would also lead to a too local model. For the same reason, for a large value of γ, we may not choose a too small value of σ.

3.7 Results

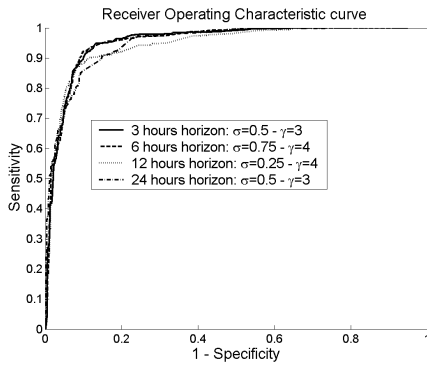Figure 5 presents the ROC curve of the optimal classifiers for 3, 6, 12 and 24 hours prediction horizons.

Fig. 5. ROC curves of LS-SVM classifiers for different prediction horizons)

Choosing a LS-SVM classifier with equal values of the two criteria NN and VV seems an interesting choice because a small improvement in one criterion would imply a big loss in the other criterion. In that case, for a 3 hours prediction horizon, the classifier is able to correctly predict ice and 'no ice' 90% of the time. For a 24 hours horizon, this accuracy drops to 85%, which remains a good result for such a long horizon.

3.8 LS-SVM classifier: conclusion

When using LS-SVM classifiers to predict the state of the road, if we choose to predict ice and 'no ice' with the same performance (which seems interesting due to the shape of the ROC Curve) we obtain an accuracy of 90% for a 3 hours horizon, and 85% for a 24 hours horizon. In the next section, we shall see that the performance can still be improved by using air temperature predictions.

4. IMPROVEMENT WITH THE USE OF A PREDICTION OF THE AIR TEMPERATURE

Up to now models only included past values of the series. In this section we study how an accurate prediction of the air temperature can help improving ice prediction.

4.1 Quality of predictions from other sources

In Belgium, predictions of the air temperature are provided among other centers by two centers (IRM, Meteoservices).

In this section we try to estimate the prediction error made by the two centers when predicting the air temperature.

The two centers provide everyday three predictions: a prediction of the air temperature of the day, of the night and of the next day.

Gathering together the meteorological stations around the places where the two centers compute predictions and comparing the mean air temperature of each group with the predictions of the two centers in Figures 6 and 7 allows us to have an idea about the accuracy of the predictions.
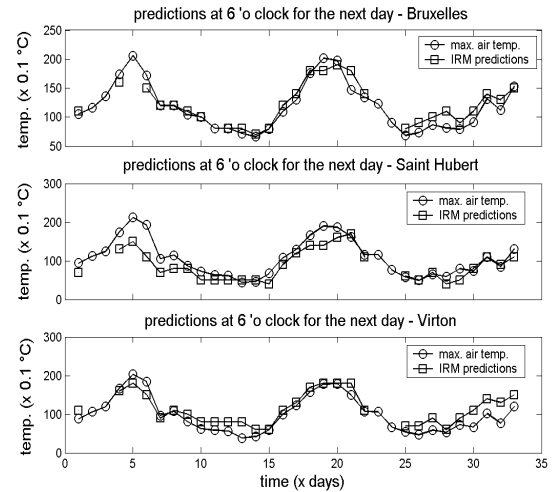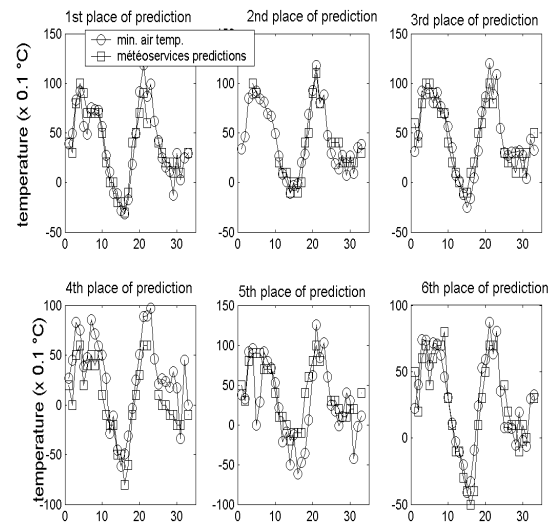

Fig. 6. Prediction of the air temperature (IRM)

Fig. 7. Prediction of the air temperature (Meteoservices)

The predictions from Meteoservices are a little better than those from IRM, probably due to the

fact that Meteoservices computes predictions in more places than IRM. Nevertheless the difference is not significative. In general the mean of the prediction error is around 0^oC and its standard deviation is less than 2^oC. Tests also showed that the prediction error of each of the two centers can be considered as a white noise.

4.2 Improvement of the LS-SVM classifier

Figure 8 shows the improvement of the ice prediction obtained by a classifier using, in addition to the inputs, a prediction of the air temperature, for 3, 6, 12 and 24 hours horizons. The mean of the error of the simulations of predictions was fixed to 0^oC and its standard deviation to 1^oC.

We observe that the use of the prediction of the air temperature allows to improve the quality of the prediction of ice.

The accuracy of the classifier improves to 95% for a 3 hours horizon, and to 91% for 6, 12 and 24 hours horizons. The use of a simulation of the air temperature allows then to divide by two the error percentage of the classifier.

5. CONCLUSION

In this paper, in order to predict the state of the road (icy or not), we tested lS-SVM classifiers. We obtained significantly better results when using a simulation of the prediction of the air temperature as one of the inputs. The accuracy of the classifier reaches around 95% for a 3 hours prediction horizon, and around 92% for 6, 12 and 24 hours prediction horizons.

Acknowledgments

REFERENCES

S. Hertl and G. Schaffar. An autonomous approach to road temperature prediction. *Meteorol. Appl.*, 5:227–238, 1998.

S. Hertl and G. Schaffar. The prediction of ice on roads. *International Journal of Modelling and Simulation*, 19: No. 1, 1999.

IRM. Irm (institut royal météorologique de belgique). *http://www.meteo.oma.be/index.html.*

Meteoservices. Meteo services. *http://www.meteoservices.be.*

K. Pelckmans, J.A.K. Suykens, T. Van Gestel, J. De Brabanter, L. Lukas, B. Hamers, B. De Moor, and J. Vandewalle. Ls-svmlab toolbox.
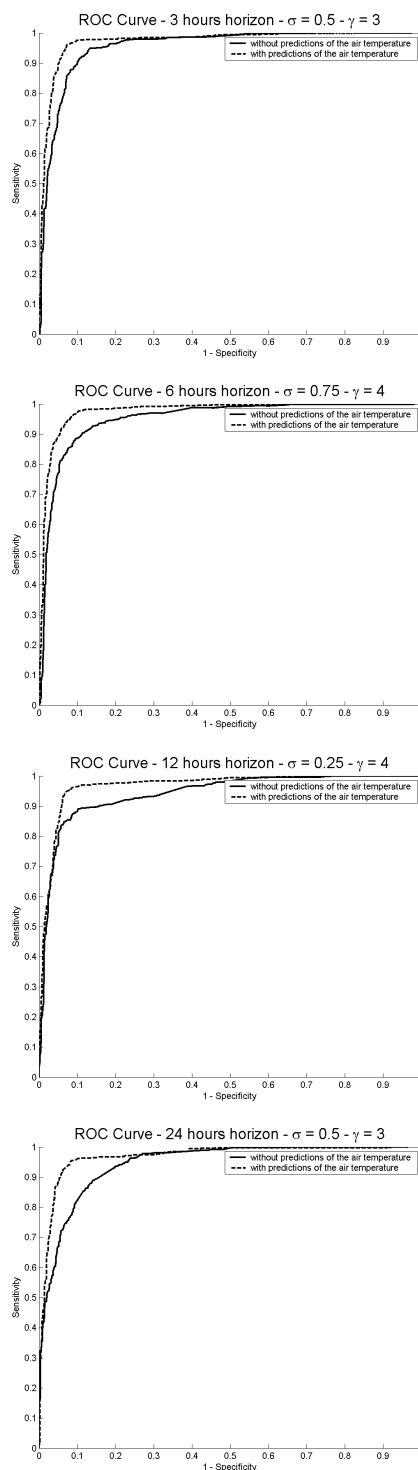
Fig. 8. Improvement of the classifier when using a IRM prediction

J. Shao. Application of an artificial neural network to improve short-term road ice forecasts. *Expert Systems with Applications*, 14(4):471–82, May 1998.

J. Shao and P.-J. Lister. An automated nowcasting model of road surface temperature and state for winter road maintenance. *Journal of Applied Meteorology*, 35(8): 1352–61, Aug. 1996.

J. A. K. Suykens and J. Vandewalle. Least squares support vector machine classifiers. *Neural Processing Letters*, 9(3):293–300, 1999.