

Using Ordinal Support Vector Machines to Model the Risks Associated with the Transportation of Hazardous Goods

J. M. Matías¹, C. Ordóñez², J. Taboada²

¹ Dep. of Statistics, University of Vigo, Spain, e-mail: jmmatias@uvigo.es

² Dep. of Environmental Management, University of Vigo, Spain.

Introduction

The transportation of hazardous goods by road implies a risk for both humans and the environment, in that an accident may cause extensive material damage and may even endanger lives. For this reason, there is a growing interest among both public and private entities (e.g. insurance companies) in studies that assess the risks associated with hazardous goods transportation.

Recently, Martínez-Alegría et al. (2003) proposed a macroscopic conceptual model for identifying the roads within a network with the greatest accident risk. This model took into account the statistical accident rate, the total and specific traffic density of vehicles carrying hazardous goods for each kind of road, the physical and chemical characteristics of the transported substances, and the vulnerability of the environmental and populational elements exposed to each kind of hazardous substance transported.

This conceptual model responded to the following mathematical model defined (U.S. Department of Transportation, 1994) for each kind of road:

$$R = G \cdot P \quad (1)$$

where R is road accident risk, G is the gravity of the potential damage and P is the probability of the adverse event occurring.

The sources of information for the above G and P terms were different, however, with G based on different factors evaluated by experts, and P based on factors whose information is obtained from historical statistical data on accidents, traffic intensity and hazardous goods transportation for roads in the network being studied.

With a view to obtaining sufficient statistical data on the factors that make up P , this model views the road as an indivisible object of the study and is, thus, fundamentally macroscopic. It is useful for the strategic analysis of a network of roads, but less so for actually managing risk i.e. for evaluating the level of risk for specific stretches of roadway and for identifying specific constructive or palliative measures for the network.

An absence of statistical data at the level of the stretch of roadway, however, obliges us to shift the focus of the model, to base it solely on factors determined

by experts. More specifically, the historical statistical factors that make up P must be replaced by new factors representative of the construction morphology of each stretch of roadway, its design and condition, as well as certain geographic accidents of the location.

Consequently, our new model was totally supervised, in other words, it was constructed from a sample of data in which both the risk inherent in a road and the factors that determine this risk were evaluated in their entirety by experts. Thus a knowledge base was created that will permit the application of the model to new regions without the intervention of experts except for the review of results.

We preferred to adopt a non-parametric focus that allowed the data to determine the most suitable model rather than an *a priori* parametric family as in Equation 1, as the latter would unjustifiably have restricted the expert knowledge model.

Thus, 33 determining factors were identified for assessing the risk of an accident occurring on a particular stretch of road. These factors were classified in two groups, as follows:

1. Accident probability factors (P): variables that affect the probability of the occurrence of an accident.
2. Environmental impact factors (G): variables that determine the potential damage to the population or the environment resulting from an accident.

Given the dimensions of the problem, the only viable statistical methods for implementing a non-linear risk model are machine learning techniques such as classification and regression trees (CARTs), multilayer perceptron networks (MLPs) and support vector machines (SVMs).

The results produced by these techniques on a test sample (Matías et al. 2004) were more favorable than those produced by traditional discriminant analysis, irrespective of any dimensionality reduction techniques used.

Following the approach developed by Herbrich et al. (2000), in this research we used ordinal support vector machines with a view to improving the results obtained. This technique takes advantage of the ordinal information contained in the data without increasing the computational load. Furthermore, the technique permits the utility function that is latent in expert knowledge to be estimated.

Definition of the model

The initial model for risk (impact) associated with an accident involving the transportation of hazardous goods along a particular stretch of roadway was constructed by combining elements of the Martínez-Alegría et al. (2003) conceptual model with specific factors that, in the opinion of the experts, affect risk at the level of the roadway stretch.

Thirty-three impact factors were identified and subsequently subdivided into two main groups, as follows:

1. Factors that affect the probability of the occurrence of an accident. This group consists of 21 factors that reflect specific features of a stretch of roadway:

- a. Design: Lane width, hard shoulder width, existence of slow lanes, types of access, existence or otherwise of protective barriers, existence and size of drainage ditches and drains.
 - b. Constructive morphology: Conditions of the road, slope, altimetry, orientation in regard to the sun, exposure to winds, etc.
 - c. Signalling and signposting: Type and condition.
 - d. Type of road works: Type of road works being performed, if appropriate, on a stretch of roadway.
 - e. Visibility threshold. Described in terms of five numeric intervals (0 to 100 m, 100 to 200m, 200 to 500 m, 500 to 1000 m, and over 1000 m).
 - f. Condition of the road: Drainage capacity of the asphalt, irregularities and defects, potholes, general state of conservation of installations, etc.
2. Environmental vulnerability factors. Adopting a conservative stance, the intrinsic danger factors that depend on the physical and chemical characteristics of a hazardous material involved in an accident were considered to be constant (i.e. we considered a single type of product classified as flammable liquid fuel), as also the extrinsic danger factors associated with the type of accident. Therefore, only vulnerability factors associated with the environment and humans were considered to be variables, with 12 of these identified as follows:
- a. Land use: Irrigated or unirrigated land, forest, nearby buildings or urban areas, industrial or mining areas, infrastructures, etc.
 - b. The natural morphology of the land: Stability of the slopes or embankments, natural slope of the land, etc.
 - c. Surface and subterranean hydrology: Distance to nearest water sources, permeability of the land and possible risk to aquifers, etc.

Our model, consequently, takes the following form:

$$R = f(X_1, \dots, X_d) \quad (2)$$

where X_i , $i = 1, \dots, d$ are the 33 factors described above, R is the risk associated with an accident on a particular stretch of roadway and f reflects expert knowledge on the influence on risk of the above-mentioned factors.

The variables X_i , $i = 1, \dots, 33$ were coded on ordinal scales of 0 to 10, with the lower values representing greater risk. Risk, in turn, was codified on an ordinal scale of 1 to 3 to indicate the hazard represented by a stretch of road (the higher the rating, the greater the number of corrective actions required).

Given the above coding system, the estimation of the model 2 can be viewed as a classification problem supervised by an expert. Our approach (Matías et al. 2004) involved the use of linear discriminant analysis, neural networks, multilayer perceptrons (MLPs) and support vector machines (SVMs). For reference purposes, results are presented below for all these methods together with the results obtained using classification trees (CART).

The above coding for R introduces a ranking of the different stretches of road in terms of suitability for transporting hazardous materials, which classification techniques ignore by minimizing the classification error rate criterion. By definition, under this criterion, two classification rules are equivalent if they result in the

same error rate. However, the classification approach fails to take account of any possible violations in the order of the examples.

With a view the improving the resulting model, we used support vector machines for ordinal data (Herbrich et al.2000), the basic concepts of which are described in the next section.

Ordinal Support Vector Machines

Assume a sample of independent observations $\{(X_i, Y_i)\}_{i=1}^n$ where $X_i \in \Omega \subset \mathbb{R}^d$, $Y_i \in \Theta$ are random variables and where $\Theta = \{r_1, \dots, r_c\}$ is a set of ordered ranks $r_i > r_j$ if $i > j$ such that $r_c \succ r_{c-1} \succ \dots \succ r_1$ where \succ is a preference relation with strict order properties (irreflexive, asymmetric and transitive).

Likewise, assume that the ranks r_i assigned by the expert are the result of a latent utility function $U : \Omega \rightarrow \mathbb{R}$, in such a way that given a point $\mathbf{x} \in \Omega$, the expert assigns rank via:

$$y(\mathbf{x}) = r_j \Leftrightarrow U(\mathbf{x}) \in [\theta(r_{j-1}), \theta(r_j)] \quad (3)$$

where $\theta(r_j) \in \mathbb{R}$, $j = 1, \dots, c$ are the values used implicitly by the expert.

Rather than a classical loss function $\ell_{0-1}(y, \hat{y}) = 1_{\{\hat{y} \neq y\}}$ that just penalizes classification errors, with a view to penalizing violations in the order produced by an ordering rule $g : \Omega \rightarrow \Theta$ with $\hat{y} = g(\mathbf{x})$, we define the following loss function (Herbrich et al. 2000):

$$\ell_{\text{pref}}(y_i, y_j, \hat{y}_i, \hat{y}_j) = \begin{cases} 1 & \text{if } y_i < y_j \text{ and not } \hat{y}_i < \hat{y}_j \\ 1 & \text{if } y_j < y_i \text{ and not } \hat{y}_j < \hat{y}_i \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

In this framework, the problem of determining the best ordering rule for the points in Ω can be viewed as a classification problem for the space $\Xi \subset \Omega \times \Omega$ of all the different pairs of points in Ω , with the label $z \in \{-1, +1\}$ defined, with $i \neq j$, as:

$$z_{ij} = z(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} +1 & \text{if } U(\mathbf{x}_i) > U(\mathbf{x}_j) \\ -1 & \text{if } U(\mathbf{x}_j) > U(\mathbf{x}_i) \end{cases} = \text{sign}(U(\mathbf{x}_i) - U(\mathbf{x}_j)) \quad (5)$$

The sample data are now: $\{(\mathbf{x}_i, \mathbf{x}_j, z_{ij}), i \neq j\}_{i,j=1}^n$.

In this context, if the expert's utility function were to apply a linear model $U(\mathbf{x}) = \mathbf{w}_e^T \mathbf{x}$, then using 5:

$$z_{ij} = \text{sign}(\mathbf{w}_e^T \mathbf{x}_i - \mathbf{w}_e^T \mathbf{x}_j) = \text{sign}(\mathbf{w}_e^T (\mathbf{x}_i - \mathbf{x}_j)) \quad (6)$$

If we resolve this classification problem using the maximum margin hyperplane following a soft-margin approach (Vapnik, 1998; Schölkopf and Smola, 2002), the problem is formulated as:

$$\min_{\mathbf{w}, \xi} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i \neq j, i, j=1}^n \xi_{ij} \right\} \quad (7)$$

$$\mathbf{w}^T (\mathbf{x}_i - \mathbf{x}_j) \geq 1 - \xi_{ij}; \quad i, j = 1, \dots, n, \quad i \neq j \quad (8)$$

Bearing in mind that the points of the sample are now difference vectors $\mathbf{v}_{ij} = \mathbf{x}_i - \mathbf{x}_j$, $i \neq j$, the solution takes the form:

$$\hat{\mathbf{w}} = \sum_{\text{s.v.}} \alpha_{ij} z_{ij} \mathbf{v}_{ij} = \sum_{\text{s.v.}} \alpha_{ij} z_{ij} (\mathbf{x}_i - \mathbf{x}_j) \quad (9)$$

where the values α_{ij} are obtained from the resolution of the dual problem in 7-8.

In the most realistic case of a non-linear utility function, we can use the kernel trick (references cited above) - which consists of transforming the data in a space with a higher dimensionality through a transformation $\phi : \Omega \rightarrow \Phi$ such that $k(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle$ is a positive definite function - and so consider the linear functions in the said space:

$$u(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x})$$

The solution in 9 is converted in this case into:

$$\hat{\mathbf{w}} = \sum_{\text{s.v.}} \alpha_{ij} z_{ij} (\phi(\mathbf{x}_i) - \phi(\mathbf{x}_j))$$

and so the resulting optimum hyperplane is:

$$\begin{aligned} f_{\hat{\mathbf{w}}}(\mathbf{x}, \mathbf{x}') &= \hat{\mathbf{w}}^T (\phi(\mathbf{x}) - \phi(\mathbf{x}')) \\ &= \sum_{\text{s.v.}} \alpha_{ij} z_{ij} (\phi(\mathbf{x}_i) - \phi(\mathbf{x}_j))^T (\phi(\mathbf{x}) - \phi(\mathbf{x}')) \\ &= \sum_{\text{s.v.}} \alpha_{ij} z_{ij} (k(\mathbf{x}_i, \mathbf{x}) - k(\mathbf{x}_i, \mathbf{x}') - k(\mathbf{x}_j, \mathbf{x}) + k(\mathbf{x}_j, \mathbf{x}')) \end{aligned}$$

an expression that means we avoid having to determine and calculate the transformation ϕ .

Consequently, the estimated utility function is:

$$\begin{aligned}\hat{U}(\mathbf{x}) &= \hat{\mathbf{w}}^T \phi(\mathbf{x}) = \sum_{s.v.} \alpha_{ij} z_{ij} (\phi(\mathbf{x}_i) - \phi(\mathbf{x}_j))^T \phi(\mathbf{x}) \\ &= \sum_{s.v.} \alpha_{ij} z_{ij} (k(\mathbf{x}_i, \mathbf{x}) - k(\mathbf{x}_j, \mathbf{x}))\end{aligned}$$

Finally, to estimate the frontiers $\theta(r_j)$, $j = 1, \dots, c$ for the intervals of the utility function that the expert implicitly uses in order to determine the ranks r_j , all we need to do is bear in mind that the pairs $(\mathbf{x}_i, \mathbf{x}_j)$ that verify $\xi_{ij} = 0$ have been classified correctly.

Therefore, if we choose a subset of pairs with ranks differing by just one unit:

$$A(s) = \{(\mathbf{x}_i, \mathbf{x}_j) : \xi_{ij} = 0, y_i = r_s, y_j = r_{s+1}\}$$

the frontiers can be estimated through the mid-point of the closest points that differ by just one unit in their ranks, in other words:

$$\hat{\theta}(r_s) = \frac{1}{2} (U(\mathbf{x}^{(1)}; \mathbf{w}) + U(\mathbf{x}^{(2)}; \mathbf{w}))$$

with:

$$(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) = \arg \min_{(\mathbf{x}, \mathbf{x}') \in A(s)} \{U(\mathbf{x}; \mathbf{w}) - U(\mathbf{x}'; \mathbf{w})\}$$

With these frontiers, the prediction of the rank that corresponds to a new point \mathbf{x} , is obtained using 3.

Estimating the model using ordinal SVM

With a view to constructing the knowledge base represented by the model in 2, 28.6 km of roadway located between the Spanish regions of Castilla-León and Galicia were selected for modeling.

This road runs through a mountainous region lying between 495 m and 1,105 m above sea-level (the deepest part of the valley and the highest part of the mountain pass, respectively). The road was constructed over twenty years ago and is very sinuous.

Two hundred and eighty-six stretches of 100 m were marked out. For each stretch the factors X_i , $i = 1, \dots, 33$ defined above were evaluated, as also the level R of corrective measures necessary for adaptation of the stretch to the transportation of hazardous goods.

Obtained as a result were 286 records of the form (X_1, \dots, X_{33}, R) . Of these, $n_{\text{train}} = 150$ were used for the estimation of the model and $n_{\text{test}} = 136$ were used as a test sample to evaluate the behavior of the different techniques.

The results are depicted in Table 1, which also includes for reference purposes, the results obtained using linear discriminant analysis, neural networks, multilayer

		Linear Disc.			Class. Tree			MLP			SVM			Ord. SVM		
Risk		L	M	H	L	M	H	L	M	H	L	M	H	L	M	H
Expert	L	17	8	0	21	4	0	19	6	0	20	5	0	24	1	0
	M	19	12	3	11	23	0	8	26	0	7	22	5	11	23	0
	H	0	6	71	0	6	71	0	6	71	0	3	74	0	6	71

Table 1. Stretches of roadway for the test sample classified according to the different models (linear discriminant analysis, classification trees, MLP, SVM and ordinal SVM) as low (L), medium (M) or high (H) risk and compared to the expert's classification (rows). The error percentages for the different techniques were 26.47% (linear discriminant analysis), 15.44% (CART), 14.71% (MLP), 14.71% (SVM) and 13.24% (ordinal SVM). The risk levels determine the extent of corrective measures required for each stretch of road in order to adapt them to the transportation of hazardous goods.

perceptron (MLPs) and support vector machines (SVMs) for classification (multi-class), as also the results obtained using classification trees (CART).

The error percentages for the different techniques were, respectively: 26.47% (linear discriminant analysis), 15.44% (CART), 14.71% (MLP), 14.71% (SVM) and 13.24% (ordinal SVM).

As can be observed, the machine learning techniques produced more satisfactory results than linear discriminant analysis, (with or without a reduction in dimensionality). The results for MLP and SVM were similar, although with a different error structure resulting from their different configurations (radial in the case of the SVM and projection in the case of MLP). Finally, the ordinal SVMs produced the best results of all.

The estimation of the intervals $[\theta(r_{j-1}), \theta(r_j)]$ for the ranks $r_j = 1, 2$ and 3 (see equation 3) were respectively: $(-\infty, -5.73]$, $[-5.73, -0.72]$, $[-0.72, \infty)$. Figure 1 shows the utility function for each of the 150 stretches of roadway represented in the training sample.

Conclusions

In this paper we have constructed a model to determine the risk of a hazardous goods accident that incorporates expert knowledge. The model can be applied on a large scale to other roads without the direct intervention of an expert, although subsequent supervision would be necessary.

To estimate the model, a support vector machine approach applied to ordinal data was compared to linear discriminant analysis and other machine learning classification techniques.

The ordinal SVMs performed more satisfactorily than the other techniques, and without increasing the computational burden to any significant extent. Moreover,

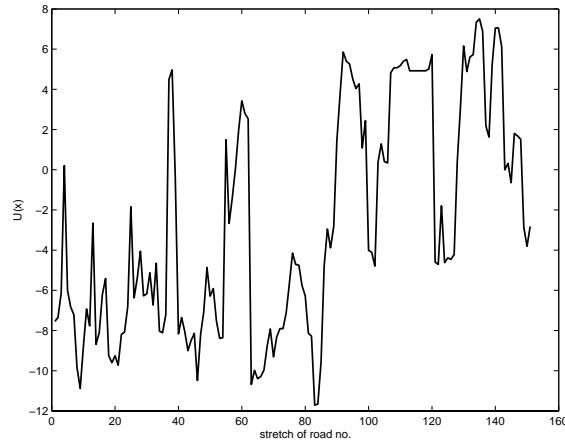


Fig. 1. Utility function for the stretches of road in the training sample numbered sequentially.

they provided an estimation of both the expert latent utility function and the decision rule used to determine the level of risk for each stretch of roadway.

The positive results would demonstrate the benefits of tackling such problems as ordinal regression problems rather than as mere classification problems, which focus on classification errors and fail to penalize inversion in the order of the examples.

References

- Fishburn PC (1985) Interval orders and interval graphs. John Wiley & Sons.
- Herbrich R, Graepel T, Obermayer K (2000) Large margin rank boundaries for ordinal regression. In: *Advances in Large Margin Classifiers*. MIT Press, pp 115-132.
- Martínez-Alegría R, Ordóñez C, Taboada J (2003) A conceptual model for analyzing the risks involved in the transportation of hazardous goods: Implementation in a geographic information system. *Human and Ecological Risk Assessment* 9: 857-873.
- Matías J, Saavedra A, Taboada J, Ordóñez C (2004) SVM and neural networks for modeling the risks involved in the transportation of hazardous goods. Technical Report Universidad de Vigo, Spain.
- Schölkopf B, Smola AJ (2002) *Learning with kernels*. The MIT Press.
- U. S. Department of Transportation (USDOT) (1994) Guidelines for applying criteria to designate routes for transportation hazardous materials. FHWA-SA-94-083. Federal Highway Administration, Washington DC, USA.
- Vapnik V (1998) *Statistical learning theory*. John Wiley & Sons.