Available online at www.sciencedirect.com
ScienceDirect
Journal of Hazardous Materials 147 (2007) 60–66
Journal of Hazardous Materials
www.elsevier.com/locate/jhazmat

Machine learning techniques applied to the determination of road suitability for the transportation of dangerous substances

J.M. Matías [a], J. Taboada [b], C. Ordóñez [b,*], P.G. Nieto [c]

[a] *Statistics Department, Vigo University, Vigo, Spain*
[b] *Natural Resources and Environmental Engineering Department, Vigo University, Vigo, Spain*
[c] *Applied Mathematics Department, Oviedo University, Oviedo, Spain*

Received 25 May 2006; received in revised form 14 December 2006; accepted 15 December 2006
Available online 21 December 2006

Abstract

This article describes a methodology to model the degree of remedial action required to make short stretches of a roadway suitable for dangerous goods transport (DGT), particularly pollutant substances, using different variables associated with the characteristics of each segment. Thirty-one factors determining the impact of an accident on a particular stretch of road were identified and subdivided into two major groups: accident probability factors and accident severity factors. Given the number of factors determining the state of a particular road segment, the only viable statistical methods for implementing the model were machine learning techniques, such as multilayer perceptron networks (MLPs), classification trees (CARTs) and support vector machines (SVMs). The results produced by these techniques on a test sample were more favourable than those produced by traditional discriminant analysis, irrespective of whether dimensionality reduction techniques were applied. The best results were obtained using SVMs specifically adapted to ordinal data. This technique takes advantage of the ordinal information contained in the data without penalising the computational load. Furthermore, the technique permits the estimation of the utility function that is latent in expert knowledge.

Keywords: Pollutant substances; Ordinal data; Machine learning; Support vector machines; Transportation

1. Introduction

The safety and efficiency of road transport is considered a strategic objective in countries like Spain, in which the proportion of goods transported by road is about 92%; 9% of road-transported goods, moreover, are classified as dangerous goods. The transportation of dangerous 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 to assess the risks associated with dangerous goods transportation (DGT). Authors such as Glickman and Erkut [1] and Cassini [2] determined risk in terms of traffic volume and population density implied by a road accident involving the release of a dangerous substance. Other authors, such as Erkut and Verter [3], Lovett et al. [4] and Fabiano et al. [5] took a different approach, and endeavoured to reduce risk by selecting alternative, lower-risk routes. Huang et al. [6] integrated GIS and genetic algorithms to evaluate the risk of hazardous materials transportation and to plan safer alternative routes. Purdy [7] analysed the risk of transporting hazardous materials by road or rail in Great Britain and concluded that the inclusion of motorist and rail passenger populations significantly affected the calculated risk levels and that the safe routing of materials with large hazard ranges may be more easily achieved by road.

Of particular interest are reports published by the US Department of Transportation, such as their Guidelines for Applying Criteria to Designate Routes for Transporting Hazardous Materials (US Department of Transportation) [8], which provide an interesting overview of advances in terms of the assessment of the level of risk associated with dangerous goods transport. More recently, Martínez-Alegría et al. [9] proposed a conceptual model for identifying the stretches of roads within a network with the greatest accident risk. These authors took into account factors such as probability of occurrence, accident type and the product transported, as also the vulnerability of the

* Corresponding author at: ETSI de Minas de Vigo, Campus Lagoas-Marcosende, 36310 Vigo, Spain. Tel.: +34 986 814052; fax: +34 986 811924.
E-mail address: cgalan@uvigo.es (C. Ordóñez).

0304-3894/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhazmat.2006.12.042

environmental and population elements exposed to each kind of hazardous substance transported. In many cases the stretches of road were over 100 km long since, given the information used to calculate accident probability, a more detailed analysis was not possible. This study can be viewed as a continuation of the study by Martinez-Alegría et al. [9]. Its aim is not to establish the level of risk for a road, which has already been studied, but rather to determine in detail – based on an analysis of 31 factors relating to the road and its physical surroundings – whether or not a stretch of road is suitable for transporting pollutant substances (data for the variables were collected at 100-m intervals).

The layout of the document is as follows: firstly, we present our model for evaluating the risk associated with an accident involving the transportation of pollutant substances. Next, we describe ordinal support vector machines (SVMs), a variation on the SVMs [10] obtained by considering a different loss function that penalises erroneous orderings. We then apply the ordinal SVMs to the problem of estimating our risk model, and compare the results to those obtained using other statistical classification techniques. Finally, we draw our conclusions on the work described.

2. Definition of the model

2.1. Impact factors

The initial model of the risk associated with an accident involving dangerous goods transportation along a particular stretch of roadway was constructed by combining elements of the Martínez-Alegría et al. [9] conceptual model with specific factors that, in the opinion of experts, affect risk on particular stretches of roadway.

Thirty-one impact factors were identified and subsequently subdivided into two main groups, namely accident probability factors and accident severity factors, discussed in turn below.

2.1.1. Accident probability factors

A total of 21 factors were considered as affecting the probability of the occurrence of an accident. These factors, which reflect the specific features of a stretch of roadway, are classified in six groups, as follows:

(a) *Design*: Road width, lane width, existence of slow lanes, types of feeder roads, protective barriers, and quality of drainage ditches and culverts. Taken as representing the lowest level of risk was a value of 7 m for road width, and a value of 1.5 m for lane width (values based on the Spanish road network). Lower values represent an increased risk, which results primarily from a reduction in the distance between vehicles and the reduced possibilities of avoiding an accident. The fact that a road has a slow lane means that faster vehicles are not necessarily affected by slower vehicles ahead (a major cause of accidents involving heavy vehicles). Feeder road types were graded in terms of a range of values, with the lowest risk associated with acceleration lanes, and the highest risk associated with direct intersections between feeder and main roads. Since protective barriers in good condition prevent animals or pedestrians from straying onto a road and causing an accident, higher values were assigned for adequate protective barriers in good condition, and lower values were assigned to protective barriers in poor condition and/or barriers that failed to fulfil their function; the lowest value was assigned when no protective barriers existed. The existence of suitably sized culverts and ditches determines the rainwater drainage capacity of a stretch of roadway, and prevents films of water accumulating on the road surface, with the resulting aquaplaning risk implied by loss of adherence.

(b) *Construction morphology*: Road condition, slope, altitude, exposure to sun and exposure to winds. Compared to wider curves, tighter curves are more likely to cause a loss of grip by a vehicle's tyres as a consequence of centrifugal forces. Slope is particularly likely to affect accident rates for vehicles travelling downhill. Long and gradual slopes place greater demands on the braking systems of heavy vehicles, with a greater likelihood of brake failure. Higher altitudes imply a harsher climate, and greater likelihood of ice and snow. Exposure to sun affects the probability of an accident, in that asphalt that does not receive direct sunlight is more likely to remain wet or to develop icy patches (which again affects tyre grip). Our area of study was located in the northern hemisphere, at latitude 40°; shadier areas were located between 330°NW and 30°NE and sunnier areas between 240°SE and 300°SW. Finally, greater exposure to wind also increases accident risk. Stretches of roadway that are exposed to strong side winds, particularly when these stretches alternate with sheltered stretches, are high accident risk areas. The risk associated with such exposed areas is determined by orthogonal orientation to prevailing winds with a west-to-east component. This factor, moreover, is aggravated by construction infrastructures, given that exposure to winds is greater on entry to and exit from bridges, viaducts and tunnels.

(c) *Signalling and signposting*: Painted road signs and lateral signs and signals. The type and condition of signalling and signposting on a road is determined above all by the field of vision in different weather conditions (rain, snow, fog, etc.), by light conditions at twilight, and by reflectivity at night. Elements that considerably reduce the risk of accidents include the existence of overhead neon-lit panels containing frequently updated information, painted road signs with good reflectivity, and legible lateral signs and signals. Moreover, road surfaces painted with anti-slip paint will also reduce accident rates. Higher values were assigned to those stretches of road without any signal and the lowest value is assigned to those stretches well signalled and painted.

(d) *Type of road works*: Existence of specific kinds of constructions on a stretch of roadway, such as earthworks, embankments, tunnels, viaducts, retaining walls, etc. The maximum value is applied if traffic flow is affected by roadworks and/or if there is only one-way traffic. This value diminishes to the minimum value (which is assigned when there are no roadworks), as roadwork bearing on traffic