Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 71 (2014) 328 – 332

Procedia Engineering

www.elsevier.com/locate/procedia

Data Mining on Fire Records of New South Wales, Sydney

Eric Wai-ming Lee[a,*], Guan-heng Yeoh[b], Morgan Cook[c], Chris Lewis[c]

[a] Deparrment of Civil and Architectural Engineering, City University of Hong Kong, Hong Kong, China
[b] School of Mechanical and Manufacturing Engineering, University of New South Wales, Kensington, NSW 2052, Australia
[c] Fire & Rescue NSW, Greenacre, NSW 2190, Australia

Abstract

This study gathered fire records from the Fire and Rescue New South Wales (F&RNSW) for investigating the most relevant event to the fire accident. Support vector machine was adopted to mimic the correlation between the information of the building and occupants and the occurrence of fire accident. The percentage of correct prediction is 65% which is considered reasonable since noise is expected to be embedded in the data of the fire records. Bayesian approach was also adopted to analyze the relevancies of the binary input parameters to the fire occurrence. Monte Carlo simulation was conducted. The result shows that the Special-Risk-Building and Smokers are the two parameters most relevant to the occurrence of fire accident.

Peer-review under responsibility of School of Engineering of Sun Yat-Sun University

Keywords: bayesian theorem, fire records, monte carlo simulation, support vector machine.

Nomenclature

A	fire accident
$\boldsymbol{b}$	constant column matri
F	either one of the binary inputs of the fire records
$\boldsymbol{r}$	vector from point $\boldsymbol{x}'$ to point $\boldsymbol{x}$
$\boldsymbol{w}$	weight matrix
$\boldsymbol{x}$	sample
$\boldsymbol{x}'$	point at the decision boundary nearest to the sample $\boldsymbol{x}$
y	label of the sample class
Greek symbols	
α_i	the i^{th} Lagrange multiplier
δ	time angle
θ	date angle
ρ	total width of the margins

1. Introduction

Most of the current fire investigations on the causes of fires rely on the practical experiences of the fire investigators. They make their judgment according to the background information of the buildings, the occupants and the evidences left in the fire scenes. Different fire investigators may arrive different conclusions due to their different experiences in fire investigations. This paper proposes to apply support vector machine (SVM) to objectively determine whether a fire is an

* Corresponding author. Tel.: +852-3442-2307; fax: +852-3442-0427.
E-mail address: ericlee@cityu.edu.hk

Peer-review under responsibility of School of Engineering of Sun Yat-Sun University
doi:10.1016/j.proeng.2014.04.047

accident or not. Also, Bayesian approach was adopted to evaluate the relevancies of different parameters to the fire accident from which precaution measures can be developed by the fire bridges to prevent the occurrences of the fire accidents. SVM was firstly developed by Vapnik[1]. It is a statistical learning model for classification task. It draws a decision boundary in the domain of the data to demarcate different groups of data by maximizing the margins between the decision boundary and the data points. The data points being used to establish the decision boundary are called support vectors. The SVM has been proven to be robust [2] in model training. Its performance was found to be superior to the traditional artificial neural network models (e.g. multilayer perceptron, radial basis function, general regression neural network, etc.).

Fire and Rescue New South Wales (F&RNSW) is the fire department in New South Wales (NSW), Sydney, Australia. They have a comprehensive system to record the fire cases occurred in NSW. This research obtained the fire records from their system to train a SVM model fordetermining the nature of a fire (i.e. accident or not) without any human intervention.

2. Brief Review on Support Vector Machine (SVM)

The SVM is a statistical learning model for classification. In the case of linear decision boundary, it draws the boundary on the sample domain to separate samples from two different classes. Assume the linear decision boundary is $\boldsymbol{w}^T\boldsymbol{x} + \boldsymbol{b} = 0$. As shown in Fig. 1 where $\boldsymbol{x}'$ is the nearest point at the decision boundary, the shortest distance between a sample x and the decision boundary is $\boldsymbol{r} = \mathrm{r}\boldsymbol{w}/|\boldsymbol{w}|$. The distance is evaluated as $\mathrm{r} = (\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} + \boldsymbol{b})/|\boldsymbol{w}|$. If we denote the points inside and outside the margin to be $y = -1$ and $y = +1$ respectively, we have the following equation.

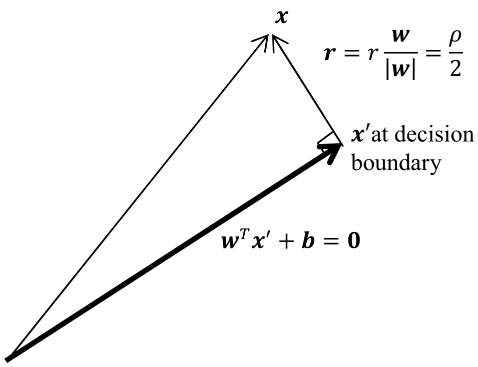

Fig. 1.Point x is a sample and $\boldsymbol{w}^T\boldsymbol{x}' + \boldsymbol{b} = 0$ is the linear decision boundary. The shortest distance from the sample to the decision boundary is denoted as margin which equals to $\rho/2$. The determination of the decision boundary is to maximize the margin.

$$\begin{cases} \boldsymbol{w}^T\boldsymbol{x} + \boldsymbol{b} \le -\frac{\rho}{2} & if\ y = -1 \\ \boldsymbol{w}^T\boldsymbol{x} + \boldsymbol{b} \ge \frac{\rho}{2} & if\ y = +1 \end{cases}$$

By combining the above two equations, we have $y(\boldsymbol{w}^T\boldsymbol{x} + \boldsymbol{b}) \ge \rho/2$. For the samples form the margins (i.e. support vectors), the equation becomes $y(\boldsymbol{w}^T\boldsymbol{x} + \boldsymbol{b}) = \rho/2$. The equation is scaled by $\rho/2$ and it becomes $\boldsymbol{w}^T\boldsymbol{x} + \boldsymbol{b} = 1$. Therefore, we have the margin $r = 1/|\boldsymbol{w}|$and $\rho = 2r = 2/|\boldsymbol{w}|$. It shows that maximizing the margin is equivalent to minimizing the value of $|\boldsymbol{w}|$or $\boldsymbol{w}^T\boldsymbol{w}/2$. The dual problem is solved by Lagrange multiplier as follows.

$$Max\left(\sum_i \alpha_i - \frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j y_i y_j\, \boldsymbol{x}_i^T\boldsymbol{x}_j\right)\text{with the following conditions.}$$

$$\begin{cases} \sum_i \alpha_i y_i = 0 & Condition\ 1 \\ \alpha_i \ge 0 \text{ for all } \alpha_i & Condition\ 2 \end{cases}$$

It results the following equations for the determination of the values of α_j.

$$\begin{cases} 1 - y_i \sum_j \alpha_j y_j \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle = 0 \\ y_i \sum_j \alpha_j y_j \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle = 1 \end{cases}$$