Corrosion Science xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Corrosion Science

journal homepage: www.elsevier.com/locate/corsci

Data mining to online galvanic current of zinc/copper Internet atmospheric corrosion monitor☆

Yanan Shi, Dongmei Fu*, Xingyu Zhou, Tao Yang, Yuanjie Zhi, Zibo Pei, Dawei Zhang, Lizhen Shao

Xueyuan Road 30, Haidian District, Beijing, China

ARTICLE INFO

Keywords:
Galvanic current
Corrosion index
Hidden Markov model
Atmospheric corrosion

ABSTRACT

The galvanic current of a zinc/copper atmospheric corrosion monitor exposed to outdoor conditions is analysed to evaluate the corrosivity of the atmospheric environment. It is essential to develop effective and efficient models for the monitored corrosion current in order to uncover the underlying mechanism of the corrosion process. In this paper, we propose a new variable, the corrosion index, to quantify the corrosivity of the atmospheric environment. The time series of galvanic current is treated statistically to predict the corrosion index via a hidden Markov model. The prediction model performs favourably on the online corrosion data in terms of efficiency and accuracy.

1. Introduction

An atmospheric corrosion monitor (ACM) composed of one insulator layer between two types of metal layers is used to monitor the real-time atmospheric corrosion behaviours of metals [1]. The long-term corrosion process is determined by the surface wetness and temperature of the two metal layers and the presence of reactive agents such as acids and salt particles in the atmosphere [2–4]. Recently, an Internet ACM (IACM) was constructed to monitor the galvanic current of the corroding metal layers and meteorological parameters and establish a connection over the Internet to transmit the necessary data in real time with high speed and high precision, meeting the requirements of online systems. With the call for sharing corrosion data [5], the innovation of IACM provides up-to-the-minute corrosion data freely available on the Internet run by the *China Gateway to Corrosion and Protection* [6] so that researchers around the world can study it at any time.

There have been some experimental studies of atmospheric corrosion related to ACM, summarised in Table 1. Most of the corrosion studies were carried out indoors in an electrolyte solution. These studies are now being gradually extended to the outdoor atmospheric environment. Different authors have obtained empirical relationships between galvanic current and environmental parameters. Detailed studies show that the conductivity of not only the electrolyte but also the salt particles is an important parameter in determining the amount of galvanic current [7,8]. The parameters that contribute to current flow, including relative humidity and electrolyte concentration, were analysed in mathematical equations [9]. Periods of dry, dew and wet conditions both outdoors and indoors could be distinguished by analysing the magnitude and time variation of the current [10]. The experiments monitoring various parts of a test vehicle demonstrated that the galvanic current of the corrosion process can be identified by driving history, temperature and relative humidity [11]. Exposure tests on carbon steel in different locations were compared and indicated that the corrosivity of the environment ranks from high to low as in marine, airport, and urban locations [12]. The ACM can be used to monitor changes in the galvanic current and environmental parameters and also to evaluate the corrosivity of a test environment [7,10,13,14].

In spite of the aforementioned outstanding works, there are still some gaps concerning the influence of environmental parameters on atmospheric corrosion. (i) The effects of some contaminants should be taken into account in the interaction of galvanic current and environmental parameters, since environmental pollution is becoming increasingly serious in recent years. (ii) It is well known that the corrosivity of the testing environment can be evaluated by ACM sensor output; nevertheless, there is an insufficient quantity of these data to measure the degree of corrosion so that the corrosivity evaluation can be converted from a qualitative to a quantitative assessment. (iii) Numerous effective representations for corrosion data mining have been proposed with artificial intelligence techniques [15–19] such as

☆ This work was supported by the Science and Technology Basic Work of Science and Technology [grant number 2012FY113000] and the Fundamental Research Funds for the Central Universities [grant number FRF-TP-16-082A1].

* Corresponding author.
E-mail address: fdm2003@163.com (D. Fu).

https://doi.org/10.1016/j.corsci.2018.02.005
Received 18 March 2017; Received in revised form 29 January 2018; Accepted 1 February 2018

Please cite this article as: Shi, Y., Corrosion Science (2018), https://doi.org/10.1016/j.corsci.2018.02.005

Table 1
Summary of some experimental studies with ACM.

Papers	Condition	Environment	Environmental parameters	Year
[7]	Indoor	Electrolyte solution	Salt particles	1976
[9]	Indoor	Electrolyte solution	RH, the electrolyte concentration	1997
[10]	Indoor, outdoor	Atmosphere	*T*, RH, time of wetness	2005
[13]	Indoor	Atmosphere	RH	2009
[12]	Outdoor	Atmosphere, marine	*T*, RH, salt particles	2010
[11]	Outdoor	Automotive	*T*, RH, salt particles, driving history	2014

T: temperature; RH: relative humidity; weather condition: SO_2, NO_2, O_3, PM2.5, PM10 and AQI (air quality index).

ANN (artificial neural network) and SVM (support vector machine). However, these effective prediction algorithms are not adopted to online data monitored by IACM due to factors such as the volatility, instability and sensitivity of time-sequential data. Unexpected failures in predicting online samples continue to be a serious problem. A deeper insight into these gaps would be very useful for a better understanding of the corrosion process.

Hidden Markov models have been used successfully in temporal pattern recognition in fields such as bioinformatics [20,21], speech [22,23] and handwriting recognition [24,25]. The basic principle is to uncover underlying information according to a set of observed data sequences, contributing to the study of the internal mechanism of the corrosion process. Moreover, HMM is a statistical model based on probability distribution, and the method can principally gain the optimal results when training samples are sufficient enough. Processing online sequential data using HMM initiates a new field of corrosion study.

In this paper, we introduce the probabilistic framework of HMM to develop a model of galvanic current in outdoor exposure based on the influence of meteorological parameters and contaminants. The rest of this paper is organized as follows. We first introduce the preliminaries of the algorithms involved in Section 2. The experiments with IACM are detailed in Section 3. The relation between galvanic current and environmental parameters are analysed in Section 4. The concept of corrosion index and experimental results are presented in Section 5. We conclude with remarks on our future work in Section 6.

2. Preliminaries

We present some preliminaries of prediction and optimization algorithms used in this paper.

2.1. Hidden Markov model (HMM)

In probability theory, an HMM is a statistical model designed to model randomly changing systems on the basis of the Markov process. The distinction between the two stochastic processes is whether the state is directly visible. We formally define the Markov process and hidden Markov model as follows [26].

Markov process The Markov process is a sequence of random variables $X_1, X_2, X_3\ldots$ with the Markov property, that is, future states depend only on the current state rather than the events that occurred before.

$$P[X_{n+1} = x|X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n] = P[X_{n+1} = x|X_n = x_n]$$

Hidden Markov model A hidden Markov model is a doubly embedded stochastic process over time with an underlying Markov process that is not directly observable, but can be observed through another set of stochastic processes that produce the sequence of observation symbols in Fig. 1.

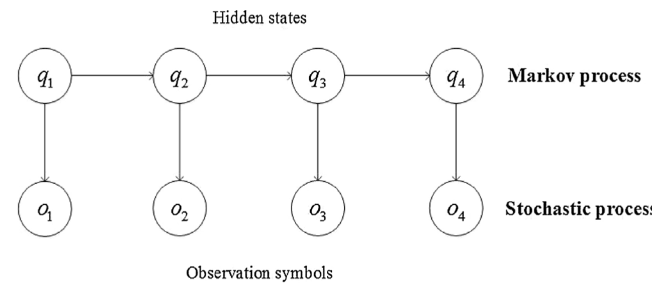

Fig. 1. A simple diagram of the HMM model.

Each hidden state in HMM has a probability distribution of the possible observation symbols. Therefore, the sequence of hidden states can be inferred statistically from the sequence of observation symbols. A basic discrete HMM discussed in this paper can be described by the following five parameters:

(1) *N*, the number of hidden states. Generally one hidden state in the model can be transferred to any other hidden state. We denote the individual states as $S = \{S_1, S_2, \ldots, S_N\}$.
(2) *M*, the number of distinct observation symbols per state. We denote the individual observation symbols as $V = \{V_1, V_2, \ldots, V_M\}$.
(3) The state transition probability distribution, $A = \{a_{ij}\}$

$$a_{\mathrm{ij}} = P[q_{t+1} = S_j|q_t = S_i] = \frac{A_{\mathrm{ij}}}{\sum_{j=1}^{N} A_{\mathrm{ij}}} \quad 1 \leqslant i, j \leqslant N \tag{1}$$

where a_{ij} is the probability that state S_i can reach state S_j in a single step. q_t is the hidden state at time *t*. A_{ij} is the number of samples for state S_i to reach state S_j. A is a matrix of $N \times N$.

(4) The observation symbol probability distribution in state *j*, $B = \{b_j(k)\}$

$$b_j(k) = P[o_{t+1} = V_k|q_t = S_j] = \frac{B_{\mathrm{jk}}}{\sum_{k=1}^{M} B_{\mathrm{jk}}} \quad 1 \leqslant j \leqslant N \quad 1 \leqslant k \leqslant M \tag{2}$$

where $b_j(k)$ is the probability that observation symbol is V_k at state S_j. o_t is the observation symbol at time *t*. B_{jk} is the number of V_k at state S_j. B is a matrix of $N \times M$.

(5) The initial state distribution, $\pi = \{\pi_i\}$

$$\pi_i = P[q_1 = S_i] \quad 1 \leqslant i \leqslant N \tag{3}$$

There are three basic problems in the practical application of hidden Markov model. One of them used in this paper is the following:

Problem – Given the observation sequence $o = o_1, o_2, \ldots, o_T$, choose a state sequence $q = q_1, q_2, \ldots, q_T$ that is optimal in some meaningful sense.

The problem listed above is a typical prediction problem. The sequence of observations gives some information about the hidden part of the model, which we can utilize to uncover the optimal hidden state sequence. The most widely used criterion is to find the single best state sequence (path), i.e., to maximize $P[q = q_1, q_2, \ldots, q_T|o = o_1, o_2, \ldots, o_T]$ [26].

Solution – Viterbi is a dynamic programming algorithm for finding the most likely sequence of hidden states. The details of the algorithm are shown in Algorithm 1.

Algorithm 1. Viterbi algorithm for HMM.

Input: a sequence of observations $o = o_1, o_2, \ldots, o_T$, hidden states *S*, observation symbols *V*, the state transition probability distribution *A*, the observation symbol probability distribution *B*, the initial state distribution π

Output: an optimal hidden state sequence $q = q_1, q_2, \ldots, q_T$