Engineering Applications of Artificial Intelligence 51 (2016) 92–96

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

On the problem of early detection of users interaction outbreaks via stochastic differential models

Markos Avlonitis

Ionian University Tsirigoti Square 7, Corfu 49100, Greece

ARTICLE INFO

Article history:
Received 21 May 2015
Received in revised form
19 November 2015
Accepted 8 January 2016
Available online 15 February 2016

Keywords:
Outbreak, stochastic differential models
Human-Computer
Interaction (HCI)

ABSTRACT

Nowadays, early detection of users' interaction outbreaks is one of the most challenging problem in Human-Computer Interaction research. In this paper a new stochastic differential model for outbreaks detection of an arbitrary time series is proposed. The proposed methodology introduces an explicit stochastic differential law for the evolution of outbreaks, mainly a limiting case of the stochastic generalization of the well known Verhustl model. The key feature of the proposed algorithm is a non-linear transformation of the corresponding population state variable. At the end, the detection of an outbreak regime in time is connected with the detection of a time regime over which the transformed variable obeys a random walk stochastic evolution. An application of the proposed algorithm is given for the problem of outbreaks in users' video interactions where the validity and usefulness of the algorithm is tested.

1. Introduction

The World Wide Web has become a very popular medium where online information due to the emergence of online social media and rich user-generated content is becoming increasingly dynamic. Human-Computer Interaction (HCI) research explores the emergence of temporal patterns of users' interactions by means of studying the association between commenting and explicitly user-generated content, in order to uncover the temporal dynamics of online content. However, uncovering patterns of temporal variation on the Web is difficult because users' interaction with online content is highly unpredictable and the corresponding temporal variation is a random process. Moreover, randomness of users' interaction is ranging within different time scales as a result of the variability of different source of contents such as news articles (Szabo and Bernardo, 2010), blog posts (Kumar et al., 2005; Liu et al., 2007), videos (Crane and Sornette, 2008), and posts in online discussion forums (Aperjis and Huberman, Wu).

The emerging scalability of temporal patterns in users' interactions has as a direct consequence the outbreaks' detection and prediction. For example, in Twitter (Java et al., 2007; Wu and Huberman, 2007), pieces of content become popular within some minutes. Online content outbreaks may emerge either as a collection of individual users interactions or as highly correlated interactions of groups of users due to the interactions between individuals (personal blogs and Twitter accounts, professional bloggers, small community-driven or professional online media sites responding quickly to events, TV stations, large newspapers, news agencies all producing content and diffusing this to other contributing users) resulting in a diffusion-like spreading of content popularity. This produces complex aggregate dynamics and patterns of users' interactions in time. Despite extensive qualitative research on the problem of modeling users interactions in time and content space (Karydis et al., 2012; Spiridonidou et al., 2013; Avlonitis et al., 2015) as well as on the emergence of temporal patterns by which content grows and fades over time (Yang and Leskovec, 2011), there has been little work about robust modeling of users' interactions outbreaks.

On the other hand, a lot of similar research has been done in other scientific areas and mainly in the field of outbreak detection of infectious diseases (Unkel et al., 2012). Indeed, a number of mathematical models have been developed since 1931 (e.g., (Shewhart, 1931)) in order to address the problem of prospective outbreak detection as they arise in a sufficiently timely fashion to enable effective control measures. These models are mainly statistical methodologies for detecting anomalies and unusual patterns in data series. More specifically, regression models generalize the Shewhart chart and are distinguished between parametric (Stroup et al., 1989; Serfling, 1963; Farrington et al., 1996) and semi-parametric (Stern and Lightfoot, 1999; Wieland et al., 2007; $author1$ et al., Zhang) based on the existence of a threshold

E-mail address: avlon@ionio.gr

http://dx.doi.org/10.1016/j.engappai.2016.01.008
0952-1976/

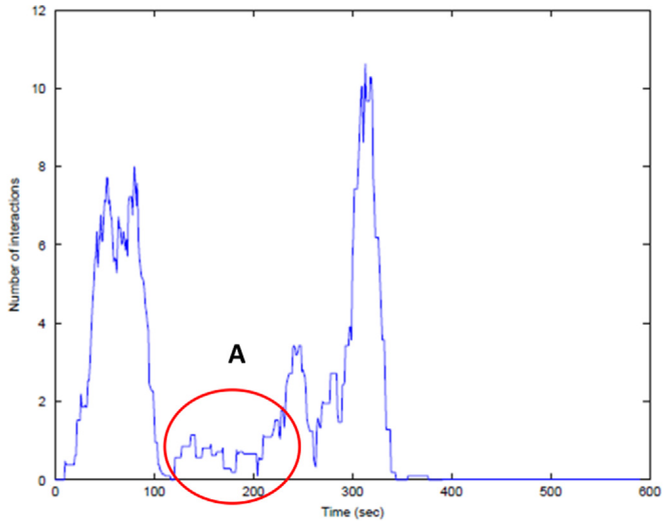

Fig. 1. The collected interaction signal: cumulative number of interactions versus relative time of the content the interaction occurred.

value whereas non-threshold methods may also be considered (Andersson et al., 2008; Bock et al., 2008; Frisén and Andersson, 2009). Moreover, time series methodologies and corresponding models have emerged in the literature where time series data were analyzed by means of the notion of the autocorrelation function (Shewhart, 1931; Williamson et al., 1999; Heisterkamp et al., 2006) as well as Bayesian and hidden Markov models (Le Strat and Carrat, 1999; Martínez-Beneito et al., 2008; Lu et al., 2010). There are also some other statistical models inspired by statistical process control as well as models accounting for spatial dynamics or for multivariate data series (Unkel et al., 2012). For the aforementioned emerging complexity of outbreak sources as well as the problem of limited data series, most of the real time applications may result to non robust outbreak estimations. Moreover, the quality of the available data, their spatiotemporal characteristics, the time scale of system evolution (e.g. explosive dynamics) as well as the appropriate selection of threshold metrics may lead to an increasing complexity. In fact, there is a generic weakness in the aforementioned statistical methods i.e., the system under study is always treated as a black box where no knowledge about the differential spatiotemporal evolution of the system is available and only the system output is treated statistically.

It is the aim of the present paper to explore the possibility of developing an explicit spatiotemporal differential class of models where the outbreak behavior can be mapped in a robust way. More explicitly, it is proposed that the well known notion of stochastic differential equations can be a powerful tool in the attempt to address the complexity emerging in real life situation of the outbreak problem. To this end, a general framework of the proposed methodology is given in the next section where the use of a specific stochastic differential model is demonstrated and analytical results are given. Moreover, in the next section, the robustness of the proposed methodology is tested in a real experiment where the outbreaks of users' interactions watching a video are predicted while in the last section some possible generalizations of the proposed methodology are discussed.

2. Stochastic differential models and outbreaks

In order to produce a robust model for the evolution of a certain system in the presence of complex internal interactions as well as interactions with a fluctuating environment the commonly

Table 1
The scheme of the Stochastic Outbreak (SO) algorithm.

Algorithm 1 The SO-algorithm
Require: Experimental time series.
for ct=L_{in} to L **do** (detection step)
r(ct) the correlation coefficient (variance function)
if r(ct) $\to 0$ No Outbreak (variance function$\to$ power law)
end if
if r(ct) > thress (variance function deviates to power law)
for w=L_{in} to ct **do** (characterization step)
transform ct to zt according to the formalism
r(zt) the correlation coefficient (variance function)
if r(zt) $\to 0$ *OUTBREAK* (variance function$\to$ power law)
end if
end for
end if
end for
return r(zt) (returns seconds of user's activity Outbreak)

encountered macroscopic systems can be described in terms of the state variables X_i obeying evolution equations of the form (Werner and Lefever, 2006)

$$\partial_t X(r,t) = f_\lambda(X(r,t)), \tag{1}$$

where $X(r,t)$ and $f_\lambda(X(r,t))$ are respectively the state variables and the functional relations expressing the local evolution of X_i's in time t and space r. The control parameter λ and the boundary conditions acting on the system constitute the constrains imposed by the external world. Moreover since the external environment is considered fluctuating the parameter λ is also fluctuating and can be interpreted as a stochastic variable. In this specific case the deterministic differential equation Eq. (1) is converted into a stochastic differential equation.

2.1. Basic aspects of stochastic differential models

In this reasoning, let us consider systems that can be modeled by a phenomenological equation of the form,

$$\dot{X} = f_\lambda(X), \tag{2}$$

where as usually, X is the state variable and λ the stochastic parameter modeling the effect of a fluctuating environment. For a wide range of applications, the functional $f_\lambda(X)$ is linear with respect to the external parameter. λ, i.e.,

$$f_\lambda(X) = h(X) + \lambda g(X), \tag{3}$$

Thus, the following equation for the parameter λ can be written,

$$\lambda_t = \lambda + \sigma\xi_t, \tag{4}$$

where for the rapid environmental fluctuations a white noise idealization ξ_t was adopted with zero mean and intensity σ^2.

Substituting in the initial evolution equation the following two different but equivalent types of stochastic differential equations (SDE), the Ito and Stratonovich SDE, results into the following expressions,

$$dX_t = [h(X_t) + \lambda g(X_t)]dt + \sigma g(X_t)dWt, \tag{5a}$$

$$dX_t = [h(X_t) + \lambda g(X_t)]dt + \sigma g(X_t)odWt, \tag{5b}$$

where the corresponding Fokker–Planck equations have the form,

$$\partial_t p_{(y,t/X)} = -\partial_y f_{\lambda(y)} p_{(y,t/X)} + \frac{\sigma^2}{2}\partial_{yy} g^2{}_{(y)} p_{(y,t/X)}, \tag{6a}$$

$$\partial_t p_{(y,t/X)} = -\partial_y \left[f\lambda_{(y)} + \frac{\sigma^2}{2} g'_{(y)} g_{(y)} \right] p_{(y,t/X)} + \frac{\sigma^2}{2}\partial_{yy} g^2{}_{(y)} p_{(y,t/X)}, \tag{6b}$$

where $dW_t = \xi_t d_t$, and o a symbol used to distinguish between a Stratonovich integral and an Ito integral (the main difference