Neurocomputing 131 (2014) 300–311

Contents lists available at ScienceDirect

Neurocomputing

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

A neurology-inspired model of web usage

Pablo E. Román [a,*], Juan D. Velásquez [b]

[a] Center of Mathematical modeling, CNRS, UMI 2807, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Av. Blanco Encalada 2120, Piso 7, P.O. Box 170/3-Correo 3, Santiago, Chile

[b] Department of Industrial Engineering, Universidad de Chile, República 701, P.O. Box 8370439, Santiago, Chile

ARTICLE INFO

Article history:
Received 18 July 2013
Received in revised form
9 October 2013
Accepted 11 October 2013
Communicated by V. Palade
Available online 29 October 2013

Keywords:
Decision making
Web usage mining
Web user
Diffusion model
Stochastic simulation
Meshless

ABSTRACT

The problem of predicting human behavior has been a great challenge for several disciplines including computer science. In particular, web user browsing behavior has been studied from the machine learning point of view, a field that has been coined web usage mining (WUM). However, current WUM techniques can be negatively impacted by changes in web site structure and content (e.g. Web 2.0). The key reason behind this issue may be that machine learning algorithms learn the observed behavior according to a particular training set, but do not model the user behavior under different conditions. We propose a simulation model that mimics human interaction with the web by recovering observed navigational steps. This web usage model is inspired by a neurophysiology's stochastic description of decision making and by the information utility of web page content. The proposed model corresponds to a high-dimensional stochastic process based on the leaky competing accumulator (LCA) neural model. We solve high-dimensional issues by considering a mesh-less symbolic interpolation. As a proof-of-concept we test the web user simulation system on an academic web site by recovering most of the observed behavior (73%). Therefore, our approach operationally describes web users that seem to react as observed users confronted by changes in the web site interface.

1. Introduction

Web usage mining (WUM) [52,55,41] corresponds to the study of the web user activities on a web site and the analysis of their information needs. For example usage profiling [53,32,31] is a technique used for such purposes. It categorizes browsing into patterns and consequently enables automatic web site customization. Recommender systems and adaptive web sites [51,53,54] are applications that benefit from WUM. Traditional data mining techniques are oriented toward applying generic machine learning algorithms to web data without considering the underlying physical mechanism. Generic analytical methods offer the advantage of rapid application, but at the cost of being limited and often requiring assistance from an expert to evaluate results. Moreover, natural phenomena described by well-grounded models are known to be much more precise for predicting behavior. Web browsing is a cognitive process that involves information foraging, visual discrimination tasks and decision making among other activities. Therefore, current cognitive science theory has been proposed using a neurocomputing point of view for modeling such tasks. We propose to describe the web user as a simple cognitive agent that is confronted by the decision of which hyperlink to click according to its own preferences. Neurocomputing is a young discipline that mixes the computational domains with the neural theory of brain processing. Solving the web user browsing behavior problem could help in both psychology and web mining. Furthermore, psychological science could benefit from web usage logs since massive visits to a web site correspond to one of the largest survey repositories of human behavior. The web mining community benefits from more accurate models of behavior. This study aims to explore this possibility by proposing an approximated model of browsing based on the neurophysiology of the phenomena. Our present work is based on an adaptation of the LCA (Leaky Competing Accumulator) model [50] for modeling web user browsing behavior. The LCA model presents a stochastic theory of decision making, and we use it [40,44] for deciding the next browsing action (e.g. clicking a link). It is conceived of as a neural-based mechanism using information accumulation and affected by random noise. At this point, we model the perceived information of each hyperlink as a random utility model [26]. This simplifying assumption enables the usage of the information scent utility for web users presented in [35], which is based on a text measure like TF-IDF [27]. Each web user is modeled as a decision maker that follows the LCA rule based on a TF-IDF vector representing web page text content, making a decision about which link to click, and repeating the process until leaving the web site. The set of artificial web users (AWU) implementing the previous model are simulated on a real web site. Each AWU is

* Corresponding author. Tel.: +56 2 29784859.
E-mail address: proman@ing.uchile.cl (P.E. Román).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.10.012

characterized by a set of parameters of the LCA and the text utility model. The simulation implements several AWU visits to the web site recording the resulting sequences of pages (sessions). The AWU's parameters represent different web session clusters that are fitted by the maximum likelihood of the time-dependent stochastic model. However, the inherent high dimensionality of the problem is an issue [5] that is proposed to be solved by using special functions and symbolic representation. Once all the parameters are fitted, the system is ready to simulate web user navigation. The rest of the paper is organized as follows. Section 2 describes related psychology of decision making. Section 3 presents the LCA model applied to web browsing and its processing. Section 4 describes the experimental setting for testing the model and procedures. Section 5 shows the numerical results of experiments. Section 6 states the main conclusions of this study. Finally, Appendix A derives the principal mathematical tools used in this study.

2. Neurocomputing of decision making

For more than 50 years, psychology has been studying the dynamic of the decision-making phenomenon. Stone in 1960 [47] stated a discrete stochastic model based on accumulation of information evidence that stops at a given threshold or within a time-out limit. A decision is made in favor of the alternative that accumulates the most information evidence. Several phenomenological studies have been inspired by the previous study using discrete or continuous diffusion models [2,21,22,12,37,8,50,9]. Recently, experimental evidence of this kind of stochastic mechanism has been performed at the neurophysiological level. Measured neural spiking rates in such cortical regions as the lateral intraparietal area (LIP) have been identified as information accumulation for visual discrimination tasks in macaque experiments [39]. Later, micro-stimulation experiments in the LIP area confirmed the assumption by observing the correct bias on final decisions of the discrimination task [14]. Recent experiments confirm that the parietal cortex processes the confidence [17] with which a decision is reached. However, several other cortex regions have been studied as information accumulators. The frontal eye field (FEF) [46] has been identified [45] as an area that performs recording, processing and motor controlling for visual discrimination tasks. The dorsolateral prefrontal cortex (DLPFC) has been identified [18] as a storage structure for spatial object location and works as an integration region. Other studies [15] based on neuro-imaging techniques suggest that decision making based on visual information corresponds to a complex set of overlapping systems working in parallel. In the case of a facial recognition task, the brain regions involved are the fusiform face area, the parahippocampal place area, the DLPFC area, the anterior insula and the inferior frontal gyrus. The mechanism of information accumulation as the neural activity level is observed in several places but mainly in the DLPFC, which fires the final decision. Stochastic diffusion processes based on information accumulation are thereby supported by the experimental fact

$$dX_i = [I_i - \kappa f(X_i) - \lambda \sum_{j \neq i} f(X_j)]dt + \sigma dW_i, \quad i = 1, \ldots, n \tag{1}$$

In this study we apply the leaky competing accumulator (LCA) model [50] in order to describe the web user decision-making mechanism. The LCA model is a multiple choice theory based on a continuous stochastic process. This corresponds to a non-linear diffusion process given by Eq. (1). Each variable X_i is interpreted as an average spike-rate activity of a set of neurons accumulating information in favor of the choice i. The system starts with no choice determination at $X_i(t=+0) = +0$, and evolves according to Eq. (1). Furthermore, as far as X_i are interpreted as biological variables [50], their values should be restricted to being positive. The decision is made in time $t = \tau$ when a variable X_{i^*} crosses a given threshold firing the decision i^*. This stopping-time process [38] results in a joint probability density $p(\tau, i^*)$ of reaching a decision for choice i^* in a time τ. However, since an analytic solution for this stochastic process does not exist, numerical approximations should be used for fitting parameters to observations. Furthermore, Eq. (1) representing a small increment dX after a small dt time interval is thereby suitable to be simulated, but only if its parameters are known. Another simple browsing model has been developed from psychological theories. The SNIF-ACT model [36] considers the web user as an agent that evaluates a given utility function of each possible browsing action on a page. It uses the ACT-R cognitive framework [1], which enables the execution of a set of production navigational rules (or browsing actions). Each of the rules is executed according to a random utility probability [35], comprising a simulation environment for web browsing. Other models like CoLiDeS [19] and MESA [30] also incorporate a browsing simulation by rules with an attention value or likelihood [20]. Furthermore, such models are not related to the mechanism for the visual perception and cognition of web pages.

3. Artificial web user model

The cognitive process of web browsing is far more complex than the model presented here. As seen in Section 2 information seeking and processing becomes an important part of the perception process. The disposition of elements on a web page may also influence the perception and motivation of each subject. Furthermore, an attention level for content does not depend only on distractions within web pages or other windows, but also in the complete environment in which the subject is immersed. Higher levels of cognition are also involved, as continuous reasoning about retrieved information could dynamically change the subject's intention memory. For example, on a web site like Wikipedia, web user reasoning would have a more important effect on the observed page sequence. Therefore, a reductionist scheme is proposed based on assumptions for perception and higher level cognition. We assume the existence of a basic set of law governing the web user browsing dynamics. As it was exposed in Section 2, the web user browsing was already described by psychology and our aim is to propose a more fundamental starting point based on the neurophysiology the phenomena. Using a set of assumption, we aims to test the web navigational theory in the simplest context. Further extension of the model could be derived to describe navigation in complex web environment, but we propose a first test for the main unperturbed phenomena as a way to state a framework for web user simulation.

3.1. Assumptions

A web site is a service for visualizing information and actions for visitors on web pages. Content is accessed and browsed by means of hypermedia facilities, in which operations correspond mainly to hyperlink clicking, back and forward buttons, URL shortcuts and direct access from search engines. Other facilities influencing web browsing behavior are the use of scroll bars, pop ups and other information disposal. Considering all those operations in a first approximation to the browsing problem would be a serious complication. For the sake of simplicity we assume:

1. Simple content disposition: Information is distributed in a few pages, and content is concisely presented on the visible portion of the page without a need to scrolling.

2. Simple content semantic: The web pages' content does not require higher order cognitive features to be understood.
3. Simple purpose web site: Web pages are created for a simple purpose, thus mainly for information disposition.
4. Content mainly based on the text: The web site presents information based on the text.
5. Single purpose visit: Web users visit the web site with a single purpose in mind.
6. No time restriction: Time pressure is not considered for the cognitive model of decision making.
7. No interruption: The browsing process is performed with full concentration on the surfing task.
8. Simple task: The purpose of the visit is simple and does not require higher levels of processing.

If information is distributed in a few short pages (assumption 1) then web operation usage like scrolling and back and forward buttons should be negligible. According to [48] 30% of all navigation action corresponds to the back button operation on a generic web site, and other similar operations are less than 2%. If relevant information is disposed on a few pages, then long sequences of pages during browsing will be rare and so will the usage of the back button. If pages can be processed in a very short time (assumption 2), it means that no higher cognitive process is involved in the dynamic of the browsing process. Furthermore, we assume that most visitors are expert users of the web site, since the service is for very precise informative purposes (assumption 3). The information is supposed to be presented using text (assumption 4), so natural language processing [27] is used for its analysis. Each hyperlink points to whatever is presented in the text (assumption 5), so web users are mainly motivated by what appears in the web page content. If we consider that the session set includes a large variety of visitors, the main statistical significance of similar visit patterns mainly refers to the decision making process. Furthermore, we assume that the process of visual perception is averaged over similar paths, uncorrelated with web user motivation, and can be considered as part of a background noise. A similar assumption was stated in [47] that considers other factors to be averaged to zero by the accumulation process. However, this assumption is no longer valid on a web site with intensive use of graphic and complex interfaces. A web site based on the text and with few options will correspond to such an assumption. Social motivation and long-term rational goals can be understood as fixed characteristics of each web user that define its current browsing behavior. Therefore, under the most simplistic assumptions the main mechanism for the browsing process is mainly influenced by mechanistic decision making.

3.2. Decision making model

We choose to study a class of models that describe neural activity evolution by a set of stochastic equations (1). The variables of the system are commonly interpreted with the averaged potential spike-count rate of groups of neurons that occur in the lateral intra-parietal area [50]. Vector $dW=[dW_1,\dots,dW_n]$ corresponds to independent Gaussian white noise. Langevin's interpretation of this equation uses the analogy with Newton's equation, where the time variation ("dX/dt" or acceleration) of the velocity (X) is proportional to the forces F that are applied. Force components F (X) are dependent on the spike rate X (velocity). The equation is interpreted as variations dX that the state of the system X experiences in a small time interval dt. The evolution is stochastic and then described equivalently [38] by a probability density function $\phi(X,t)$ in a domain Ω (Fig. 1).

The dynamics of the stochastic variable $X(t)$ are also affected by a border condition and an initial distribution. Our model considers

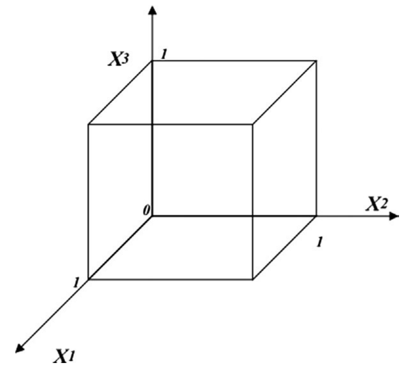


Fig. 1. A domain Ω for variable X with $a_i=0$ and $b_i=1$ for $i=1,2,3$.

starting the process at the origin $X(t=0)=0$, in which case the initial probability distribution is given by Dirac's delta function (2). The variable domain is a box $\Omega=\{X=[X_1,\dots,_n]|\ a_i\le X_i\le b_i\}$ that should contain 0. The process after a while will eventually reach the border of the box and remain confined to it (see Fig. 1). Two kinds of border condition are established, absorbent (Eq. (4)) and reflecting. Absorbent boundaries ($\Psi\subset\partial\Omega$) are related to process termination reaching a decision determination. In such cases, the stochastic process corresponds to a stopping-time problem. An approach to finding the stopping-time probability corresponds to using the complementary problem of never reaching the halting condition, in which case corresponding boundary constraints should be absorbed. The result is that the probability density of no decision $\phi(X,t)$ should tend toward zero as the absorbing boundary Ψ is approached (see Eq. (4)). Therefore, the stopping-time condition is now stated as a boundary condition. The probability $P(T)$ of reaching a decision in time $t=T$ will correspond to $P(T)=1-\int_\Omega\phi(X,t=T)\,dX$.

$$\phi(X=x,t=0)=\delta(x) \tag{2}$$

$$0\le X_i(t)\le 1,\quad (a_i=0,b_i=1),\quad \forall i,t>0 \tag{3}$$

$$\phi(X\in\Psi,t)=0,\quad \Psi\subset\partial\Omega,\quad \forall t \tag{4}$$

$$J(X,t)\cdot n(X)|_{X\in\Delta}=0,\quad \Delta\subset\partial\Omega,\quad n\perp\Delta,\quad \forall t \tag{5}$$

The reflective boundary ($\Delta\subset\partial\Omega$) constrains the system's further evolution to the box Ω, if variable X cannot cross the boundary Δ, then the orthogonal component of the probability flux $J(X,t)$ must vanish (Eq. (5)). The vector $J(X,t)$ will be defined later using the Fokker–Planck equation (12) for the density $\phi(X,t)$. The I values are the forces that drive those equations (1). Furthermore, we consider that those values are proportional to the probability $P(i)$ of the discrete choices ($I_i=\beta P(i)$), which are usually modeled using the Random Utility Model (RUM) [49]. Discrete choice preferences have been studied in economics to describe the amount of demand for discrete goods where consumers are considered rational as utility maximizers. In this study we apply this concept to quantify the subject's willingness to follow a hyperlink [24,25].

$$F=I-\omega X,\quad \omega=[\omega_{ij}] \tag{6}$$

$$\Psi=\{X=[X_i]\in\Omega|\exists j\ X_j=1\} \tag{7}$$

$$\omega_{ij}=\lambda\delta_{ij}+\kappa(1-\delta_{ij})\ \forall i,j \tag{8}$$

The utility maximization problem regarding discrete random variables results in a class of extreme probability distributions, in which the widely used model is the Logit model (Eq. (9)) and where probabilities are adjusted using the known logistics regression [33]. The Logit probability distribution $P(i)$ anticipates every possible choice on the page j and has a consumer utility V_j.

$$P(i) = \frac{e^{V_i}}{\sum_{j \in C} e^{V_j}} \tag{9}$$

The utility function is proposed to be dependent on the text present in links that the agent interprets and through which it makes the decision. Hence the assumption is that each agent's link preference is defined by its TF-IDF text vector μ [27]. The TF-IDF weight μ_k component represents the importance to the web user of the word k. Furthermore, an agent prefers to follow similar links to its vector μ. The utility values (Eq. (10)) are given by the dot product between the normalized TF-IDF vectors μ and L_i that represents the TD-IDF weight text vector associated with the link i.

$$V_i(\mu) = \frac{\mu \cdot L_i}{|\mu \| L_i|} \tag{10}$$

The resulting stochastic model (Eq. (1)) is dependent on the parameters $\{\kappa, \lambda, \sigma, \beta, \mu\}$ and the set of vectors $\{L_i\}$. The μ vector must be considered as an intrinsic characteristic of each web user that motivates them to seek some kind of content. Parameters L_i are input from web content, and the other $\kappa, \lambda, \sigma, \beta$ are considered related to the neurophysiology of the subject. The possibility of refining the model, by adding new force terms for better adjustment of the theory to observations, is a main tool in the quest for describing the nature. Furthermore, numeric manipulation can take advantage of the force abstraction, and the boundary condition could be expressed in the stochastic equation by means of fictitious forces. A reflective condition can be approximated by a force field that is negligible in the interior of the domain Ω and strong on the reflective border directed into the interior of the box. An absorbing boundary corresponds to a never-returning surface, thus forces are null in the interior and stronger on the border pointing outside the box. In one dimension this absorbing force is given by $F^A = (1-X)^{2N}$ with a large integer N. Indeed the reflective boundary in one dimension can be replaced by a large positive force for $X < 0$, and nearly zero for $X > 0$. In such a case a more formal representation of the reflective force is given by $F^R(X) = X^{2N}$, with a large integer N. In Fig. 2 the boundary force $F^R + F^A = X^{2N} + (1-X)^{2N}$ is presented for several values of N. This force has a polynomial form, whose advantage is to help in calculations involving integrals, since even in the multivariate case integrals can be calculated exactly. Such a force in the multivariate case in Eq. (11) replaces the boundary condition as an additive term, and on the limit when $N \to \infty$.

$$F_i^B = X_i^{2N} + (1-X_i)^{2N} \tag{11}$$

Therefore, numerical methods involving solving partial differential equations require smooth initial and border conditions, so that this approximation with $N \to \infty$ introduces an easier way to solve the problem. Initial conditions (2) become the only restriction that needs to be imposed. The stochastic equation (1) can be solved in terms of the probability-density function $\phi(X = x, t)$ of being in the middle of the decision-making process at time t. There are two cases depending upon whether the condition for choice determination depends on crossing a threshold or not. The probability density of never reaching a decision implies considering $\phi(X \in \Psi, t) = 0 \forall t$, corresponding to an absorbing condition. Therefore, the density $\phi(X, t)$ is described by a partial differential problem called the Fokker–Planck [10] or the forward Kolmogorov equation [38]. The forward version was chosen for its continuity equation interpretation and the use of fictitious forces for approximating an absorbing border condition, which terminates the process.

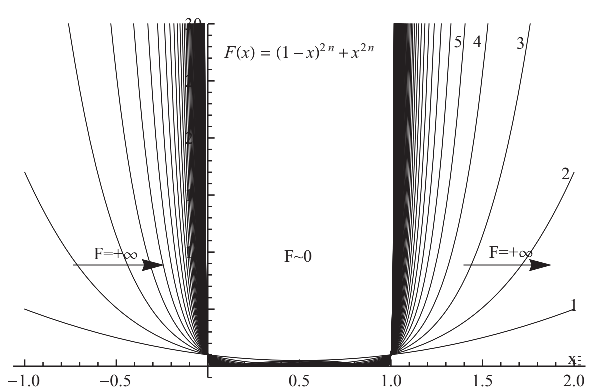


Fig. 2. Boundary forces in 1-D.

$$\frac{\partial \phi(X,t)}{\partial t} = \nabla \cdot [-\phi(X,t)F(X) + \frac{\sigma^2}{2}\nabla\phi(X,t)] = L\phi \tag{12}$$

Eq. (12) describes the evolution in time of the density of being in the middle of the decision-making process at time t. The dynamic is now in the form of a Cauchy problem $\partial\phi/\partial t = L\phi$ with the initial condition $\phi(X, t=0) = \phi_0(X)$. Operator L is linear in ϕ and based on derivatives on variable X, meaning that any time-derivative of ϕ can be obtained by a power of L. If the initial condition in $t=0$ is known, it can be propagated over time because any derivative is known (Eq. (13)). This is the basis of the semi-group property of time translation for density $\phi(X, t)$. Notice that border condition could be included in the operator L by means of the force F just by adding a fictitious term to it ($\bar{L} = L + F^B$). The approximation will be better as F^B simulates border conditions. It will be shown later that reflective conditions are easier to enforce approximately in the solution, but it is not the case for absorbing conditions. In such cases, the propagation operator will include only the F^A force:

$$\phi(X,t) \simeq e^{t\bar{L}}\phi_0(X) \tag{13}$$

As seen, the probability $P(T)$ of reaching a decision in time $t=T$ is the complement of the previous probability $P(t) = 1 - P_0(t)$. The distribution density of reaching a decision $p(t)$ is given by the derivative of $P(t)$.The probability density of choosing i at time t is derived in the appendix for Eq. (A.28) and corresponds to the following expression:

$$p(i,t) = -\frac{\sigma^2}{2}\int_0^1 \cdots \int_0^1 \frac{\partial \phi}{\partial X_i}\Big|_{X_i=1} \prod_{k \neq i} dX_k \tag{14}$$

Eq. (12) has a well-known solution for the free case, i.e. without considering the border condition. In such cases, the model is called the Ornstein–Uhlenbeck (OU) process [34]. An exact solution of the OU process is given in [10] which is revised in A.2 and in matrix notation in Eqs. (15) and (16). The importance of having such a solution is to have a basis for constructing an approximation to the boundary problem:

$$\phi(X,t) = (2\pi)^{-n/2}|\Sigma(t)|^{-1} e^{(X-It)^T \Sigma(t)(X-It)} \tag{15}$$

$$\Sigma(t) = \frac{1}{\sigma^2(e^{2\omega t} - 1)\omega} \tag{16}$$

The perpendicular gradient $\nabla\phi \cdot \hat{n}$ of a solution can be nullified on reflective border Δ by means of the technique of reflecting images (e.g. [38]. The perpendicular gradient as part of the flux vector (Eq. (A.29)) plays an important role on the reflective condition (5). If the gradient vanishes then a term whose magnitude is ϕF remains that quickly drops to 0 as the function evolves due to reflective forces. (Note that F remains bounded in Δ.) As shown in A.4 the reflective operator $U_n[.]$ does the work for any dimensions n of the space Ω.

3.3. Fitting procedure

Observations are not performed on subjects with the same purposes, which are mainly represented by vector μ. Recorded sessions represent a distribution of subjects according to a set of intentions. Using a kind of representative value $\bar{I}$ as a partial solution could be a way for finding the calibration of the model. Maximum likelihood is a well-known technique for stochastic model calibration. The likelihood of observing the probable available data is maximized using variables as the unknown parameters, subject to the restrictions of the theoretical model. The solution is interpreted to be optimal in the sense of being the most probable according to the observation. The calculation for obtaining a kind of average $\bar{I}$ vector is based on the observed data. For a given possible choice i, a distribution (in k) of the number n_{ik} of observed selections measures the time spent for deciding t_{ik}. In this context the log-likelihood is given by Eq. (17), where $p(i,t|I)$ is the probability of choice in Eq. (A.34) (expanded in (18)) for a given value of I. The value of (18) is positive since $\phi \geq 0$ on Ω and $\phi = 0$ on Ψ, in which case the derivatives are negative. The constraint stated in (19) considers $\beta = 1$ and I to be a probability:

$$\max_{I,\kappa,\lambda,\sigma,\phi} S = \sum_{ik} n_{ik} \log(p(i, t_{ik}|I)) \tag{17}$$

$$p(i,t|I) = -\frac{\sigma^2}{2} \int_0^1 \cdots \int_0^1 \frac{\partial \phi}{\partial X_j}\Big|_{X_j=1} \prod_{k \neq j} dX_k \tag{18}$$

$$\sum_k I_k = 1, \quad I_k > 0, \quad \sigma > 0, \quad \kappa > 0, \quad \lambda > 0 \tag{19}$$

Considering the probability density spanned by polynomial function to a degree of approximation of D, the power of variables X_i drops to 0 quickly with D since variables $X_i \in (0,1)$ have positive values less than 1

$$\phi(X,t) \sim \sum_{d=0}^{D} \sum_{[\sum_i k_i = d]} a_k e^{-\alpha_k t} \prod_{j=1}^{n} H_{k_j}\left(\frac{\sqrt{D_{jj}}[RX]_j}{\sigma} - \frac{[RI]_j}{\sigma\sqrt{D_{jj}}}\right) \tag{20}$$

The previous expression relies on a combinatorial partition of the integer d represented by the integer vector $k = [k_1, \ldots, k_n]$. Therefore, probability can be calculated by replacing in (18) resulting in the following equations:

$$p(i,j,t|I) \sim -\frac{\sigma^2}{2} \sum_{d=0}^{D} \sum_{[\sum_i k_i = d]} a_k e^{-\alpha_k t} \frac{\partial}{\partial X_j} [S_{k_l}(j, X_j, D, I, \sigma)]_{X_j=1} \tag{21}$$

$$S_{k_l}(j, X_j, D, I, \sigma) = \int_0^1 \cdots \int_0^1 \left(\prod_{l=1}^{n} H_{k_l}\left(\frac{\sqrt{D_{ll}}[RX]_l}{\sigma} - \frac{[RI]_l}{\sigma\sqrt{D_{ll}}} \right) \right) \prod_{k \neq j} dX_k \tag{22}$$

Despite the presence of a polynomial on the integral (22) it is difficult to obtain an explicit expression for $S_{k_l}(j, X_j, D, I, \sigma)$. It turns out much more unmanageable to calculate this integral by using a partition on the set Ψ. Indeed, the number of points in the grid is exponential over the number of dimensions of Ψ. Symbolic integration (e.g. [7]) can manage very efficiently the computation of $S_{k_l}(j, X_j, D, I, \sigma)$. A first observation is that the multivariate integrated function is a polynomial on variables $\{X, I, a\}$ and a rational function on $\{D, \sigma\}$. Therefore, integrals over variable X are straightforwardly calculated and evaluated by symbolic integration. This observation is important since it drastically reduces the computational complexity of the inference algorithm. Furthermore, derivatives of $S_{k_l}(j, X_j, D, I, \sigma)$ on I, σ, κ, and λ can be directly extracted after symbolic processing. In this way, traditional nonlinear optimization methods can be used on this system using the resulting evaluation of the function S. Symbolic processing uses fictitious forces (11) for emulating the border conditions. Such forces are polynomial functions on variable X when the propagator operator (Eq. (23)) can be approximated to any order in t symbolically. Such an operator performs the evolution of a system that fulfills an equation $\partial\phi/\partial t = L\phi$ according to Eq. (24)

$$U(t) = e^{tL} \tag{23}$$

$$U(t')\phi(X,t) = \phi(X, t+t') \tag{24}$$

Approximating boundary conditions and using the propagator operator generates a symbolic solution for a time t from polynomial approximation. The likelihood problem does not depend on the a_k parameter being simpler to implement, since it does not depend on a discretization scheme over a high-dimensional space. This is a very important observation. The number of variables $|a_k|$ is nearly exponential on the number of partitions of d (degree of the polynomial). For example, if the number of variable X is 20 and the maximal degree of the polynomial is 8 then the number of variable $|\{a_k\}|$ is 3, 108, 105. This fact introduces a limit into the degree of the polynomial versus the available computational capabilities. Nevertheless, once an initial solution is selected, it will consist of a polynomial objective function of degree equal to d and with a number of variables $|\{I_i\}|$. An approximation of the initial solution can be realized with the solution (A.26) transformed with the linear operator defined in (A.36), this procedure ensuring a solution that fulfills the perpendicular gradient on Δ and on $t = \varepsilon \sim +0$ accumulated near $X = 0$. Variables $X_i \in (0,1)$ ensure that the power tends toward 0 as the exponent increases, after which the polynomial approximation of such a solution makes sense. In this sense the exponential can be replaced by a truncated Taylor series to the degree D. In this case the Hermite polynomials have a linear relation to the monomials in variables X_i. A matrix-inversion problem is stated for finding variables a_k in order that ϕ approximates a solution (A.26) to degree D. This notably simplifies the application of the operator e^{tL} since now ϕ is spanned by means of the sum of eigenfunctions of the L operator resulting in the exponential of the series. It is important to consider the normalization of this function on the initial state. It is important to notice that the number of reflections on a multivariate set is exponential, so a careless implementation of the operator (A.36) could result in more computer time. However, there is a simple solution using an iterative method just as in the induction demonstration, by applying one reflection by coordinate each time. The resulting algorithm executes in a number of steps that depend linearly on the number of dimensions. This approximation has the cost of not exactly accomplishing the border condition. The part corresponding to the reflective border Ψ will contribute as an additional probability mass to the total distribution. However, the part corresponding to the absorbing border will quickly vanish since it is on the border. Also as we consider the solution (A.26) in $t = \varepsilon \sim +0$ the extra mass on the absorbent border will be negligible as $t \to 0$. A more important error source corresponds to the positive contribution that ϕ adds to the orthogonal flux on the border Δ. Evolution on the domain Ω with the propagator operator e^{tL} implies that any mass near or on the border Δ will go to zero as time passes. Nevertheless, such an initial non-compliance of reflective border condition implies an additional mass effect for the total probability. A multiplicative parameter on ϕ should help to adjust the difference.

3.4. Simulation procedure

Monte Carlo techniques become part of the standard treatment of a high-dimensional problem. Nevertheless, even such methods lose a high degree of accuracy value when a high number of dimensions are involved. More accurate methods like quasi-Monte Carlo techniques [23] and exact simulation [13] are envisaged for embracing the computational complexity. The stated diffusion model of the decision-making class is suitable for simulation by exact simulation due to its specific properties. Simulation is used mainly as a side-validation technique for obtaining numeric bounds for the accuracy value of the calibration and for experimenting with the conditions of the model. Exact simulation (e.g. [3,4]) uses Ito-integrated stochastic equations (e.g. [34]) for simulation by means of small time steps. Hopefully the transformation R (A.1) has the following property $R=R^{\dagger}=R^{-1}$ that implies that the transformed random white noise RdW continues to be a random white noise vector. The R transformation is detailed in the (A.1) and can be applied over the stochastic equation (1). Furthermore, simulation for a stochastic equation (1) can be derived from exact simulation of the OU process including border conditions. For a point in the interior of Ω, whose time increment is sufficiently small to not reach the boundaries, then the stochastic evolution of the vector X is given by (25). The solution is obtained by common Ito stochastic integration [34].

$$X(t)=e^{-\omega t}X(0)+M(\omega,t)I+\sigma K(w,t)Z \tag{25}$$

The random vector Z has each component behaving as normal variables $Z_i \sim N(0,1)$. Such a vector equation includes a matrix ω whose components are $\omega_{ij}=(\kappa-\beta)\delta_{ij}+\beta$. ω which can be diagonalized $\omega=RDR$ using the orthogonal transformation $R_{ij}=(1/\sqrt{n})[\cos(2\pi kj/n)+\sin(2\pi kj/n)]$, where $[D]_{ij}=(\kappa-\beta)\delta_{ij}$ if $i\neq n$ and $D_{nj}=(\kappa+(n-1)\beta)\delta_{nj}$. Functions over the matrix ω are defined over each diagonal element as $F(\omega)=RF(D)R$, where $[F(D)]_{ij}=F(D_{ii})\delta_{ij}$. Furthermore, the exponential function on (25) is an nxn square matrix defined as before and the rest of the terms are described

$$M(\omega,t)=(1-e^{-\omega t})\omega^{-1} \tag{26}$$

$$K(\omega,t)=[\tfrac{1}{2}(1-e^{-2\omega t})\omega^{-1}]^{1/2} \tag{27}$$

Such a term reflects consistency on $t=0$ with Eq. (25), since $M(\omega,0)=0$ and $K(\omega,0)=0$. Despite the apparent complexity of matrix function, processing such matrices by computer program is straightforward. Indeed they depend on the matrix ω, which is diagonalizable as seen in (A.1). Therefore, any C^{∞} function of ω is computed on each eigenvalue of the diagonal form and transformed back using R. An efficient and accurate algorithm for simulating the web user is described in Algorithm 1.

Algorithm 1. Simulation of the stochastic equation (1).

Data: Vector $X=[X_i]$ is initialized near +0 and t,h with a small value. Vector $I=[I_i]$ is calculated from web page content.
Result: Return the next hyperlink i^* that the AWU will follow.
while *$X(t+kh)$ is not close to the absorbing Ψ* **do**
  Evaluation : $e^{-\omega(t+kh)}$, $M(\omega,t+kh)$ and $K(\omega,t+kh)$ according to (26)and(27);
  Exact simulation : $X(t+kh)=e^{-\omega(t+kh)}X(t)+M(\omega,t+kh)I+\sigma K(w,t+kh)Z$, where Z is a generated vector of normal $N(0,1)$components;
  if *$X(t+kh)$ is close to reflective Δ* **then**
    Set $X(t+kh)=X^*$ to be the point on the straight line on the border of the neighbor of Δ;
  end
  Set $k=k+1$;
end
Return the decision time is $\tau=t+kh$ and the decision reached is $i^*=\text{ArgMax}\{X_i\}$;

stated in Section 3.1. It mainly presents the department of industrial engineering and academic programs, whose content is mainly based on the text. Web users who visit this web site have simple motivations such as looking for specific academic program information. Web users are considered stochastic agents [42,43]. Those agents are supposed to follow the LCA stochastic model dynamics (Eq. (1)), and maintain an internal state X_i with some white noise dW_i. The available choices lie in the links on a web page, including the probability of leaving the web site. Agents make decisions according to their internal preferences using a utilitarian scheme. Collected data from a real web server contains the behavior of a variety of different users. A parameter inference should be performed on the distribution of the μ vectors to discover the web user's preferences. A web user is considered memoryless, making decisions without considering the previous pages visited, but having a purpose driven by μ. A special link corresponding to the decision of leaving the web site is presented on every page, with a fixed probability transition analogous to the random surfer teleportation operation [6]. Each artificial user ends up following a trail $((p_1,t_1),\dots,(p_L,t_L))$ of pages $\{p_o\}$ with the visitor's time durations on the site $\{t_o\}$, until the moment the user decides to leave the L step. As was mentioned, the parameters of the decision model are separated into two, the evidence vector (I) and the neural tissue constant $(\kappa,\lambda,\beta,\sigma)$. An assumption was performed considering the simulation for a fixed evidence vector, built on the base of a μ preference text vector that contains the web site's most important word. This was achieved by cutting the higher 10% of the TF-IDF terms and the lower 5% of the values of the whole site. The resulting simpler stochastic model has only four scalar parameters. In this research we used five sub-sites belonging to the Industrial Engineering Department. The main departmental site, three sub-sites from the masters degree program, and a project web site contain nearly 1000 web pages. Each one has its own characteristics in terms of content and structure and there is no homogeneity in relation to the process of web construction. Only one uses a content management system which allows for standardizing the addition of content, but in the others the insertion is manual. The main topics addressed on these web sites include general information about the Industrial Engineering Department, faculty, staff, descriptions of the undergraduate and post-graduate programs and news and information about upcoming events and conferences, among others. This web site has the characteristic of having a lower degree of complexity and minimal changes to the web site when compared with others. Those characteristics make sessions simpler and ideal for the study of WUM. Real sessions are retrieved and stored in the web log format

4. Experimental setting

As a “proof-of-concept”, we choose our academic web site (http://www.dii.uchile.cl). It fulfils most of the requirements in order to test sessionization. The web site contains nearly 17,000 stemmed terms. Furthermore, the content is much more dynamic than the structure. Changes in the content reach on average nearly 347,000 changes in term frequency per month. This corresponds to the updates that are performed on daily news over the entire site. Each page has on average 193 different terms. It was decided to use the WebSPHINX Java library for implementing a crawler. This system periodically (every day at 6:00 and 12:00P.M.) recursively inspects all pages from the web site, extracting hyperlinks to other pages and the text content. Text was filtered eliminating stop words and realizing a stemming process [27]. As a result of the processing of one page the following data is obtained. A Unix timestamp corresponding to a unique time value of the retrieval of the next object. A URL and page Title serves as unique identifier of the web page. A set of hyperlinks gives the structure of the web site. Finally, a set of stemmed words corresponding to a list of each term was found on the page with its corresponding number of appearances. The number of hyperlinks changes over a month. The average number of hyperlinks on the web site is 4058 with an average change per month of −109 (2.7%) which means that the number of hyperlinks was decreasing. The number of pages also changes, with an average per month of 691 different pages with a change of −15 pages per month (2.2%). Those pages correspond to the sub-sites where the agreement allows for investigating web user sessions. All those data are stored in a relational database. This storage works with the following idea that there exist permanent pieces of information that are stored only once. Such data are keywords that at some moment will reach more than 20,000 terms but after that the growth of the table will be negligible. The same phenomenon occurs with pages represented by a URL and hyperlink. On the other hand, other attributes like the number of appearances and the time validity change periodically. Such dynamic properties are stored on tables with the postfix "detection". In this way it was possible to store the whole history of changes on the web site.

5. Experimental results

The calibration was verified using a Monte Carlo simulation iteration where changes in parameter were tested and best matching with the time distribution of sessions was used. The following results were obtained:

1. λ: 0.4
2. κ: 0.2
3. σ: 0.03

The distribution obtained matched the asymptote linear parameter (Fig. 3) within 10% of error. Those parameters were then fixed for performing the calibration of the evidence vector. The symbolic-based algorithm was stated and a single vector was fitted. An interesting fact is that I is very similar to the vector obtained by the most important word on the web site. A 40% error rate was identified in the session distribution simulation. The process took nearly 10 h. A vector distribution was obtained after clustering the session by using the longest common subsequence distance for clustering. The clustering process was stopped when 10 clusters were detected. The same process of calibration as before was performed using this method on each cluster and subset of visited pages. The results were astonishing, in that nearly 8.3% of error in the simulated session distribution was obtained. As we have seen, session length follows a typical distribution [16]. The simulated "average web user" follows a distribution of session length similar to the empirical distribution. The relative error is 8.1%, less than 1% of error in log scale. Distribution error remains more or less constant for sessions with a capacity of less than 15 pages consisting of 0.3% of error (Fig. 4). It is not surprising that the leaving-site probability chosen was equal to the sum that was

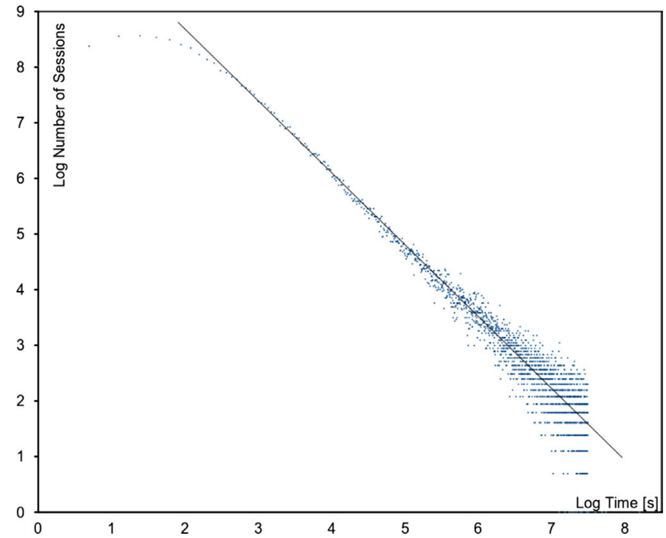


Fig. 3. The distribution of time duration of a session in log–log scale.

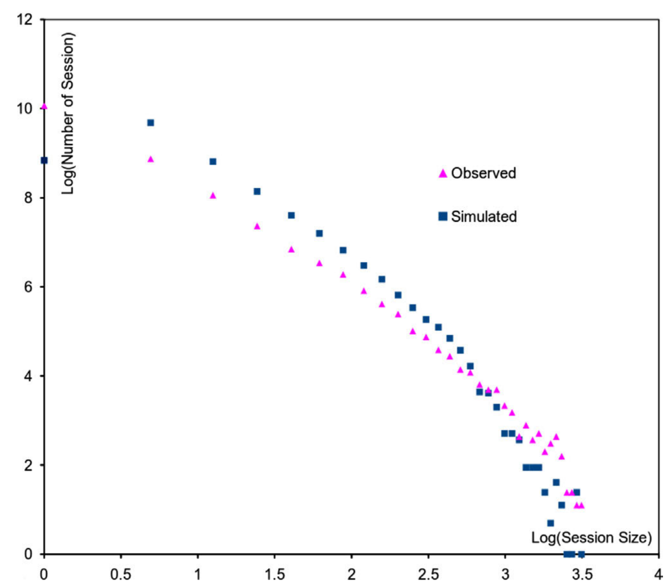


Fig. 4. The empirical (squares) vs. simulated (triangles) distribution of session length in log scale.

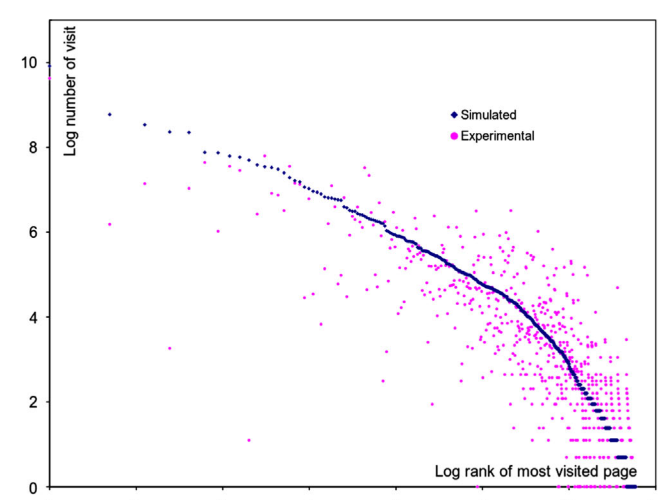


Fig. 5. The distribution of visits per page on simulated vs. experimental sessions.

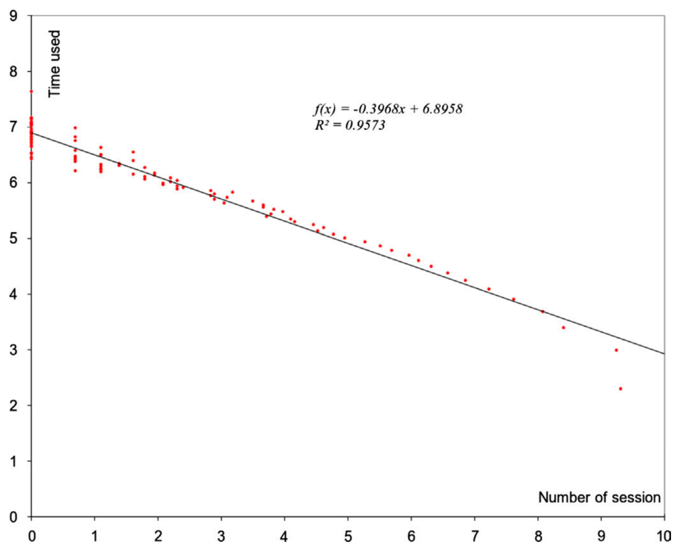


Fig. 6. The ranked time spent on a session in log scale.

empirically calculated. However, several other parameter variations were carried out that did not succumb to such degrees of adjustment. Therefore, this is a relevant result. Fig. 5 compares the visited page frequencies of the simulated behavior with the experimental behavior, with nearly 5% of error and 50% of variance. The time behavior of the model is revealed to have the same power law shape as with real sessions (Fig. 6). The regression results obtained from the distribution of the μ text-preference vector show that most probably (maximal $P(u)$) have the following word highly ranked. Three vectors were selected within the upper 70% of probability. A sample set of the words obtained by this method is presented:

1. *Management, Engineering, Enterprise, Society*. Interpretation: related to the industrial engineering field.
2. *Mgpp, Council, Bank, Description. Interpretation*: Word related to a master's degree in public politics (Mgpp).
3. *Capital, Market, Sustainability, Characterization*. Interpretation: Word related to economics.

Those results show great accuracy in describing the interest of real visitors to the web site of the Industrial Engineering Department of the University of Chile. It was known that traditional machine learning results in 50% effectiveness for rebuilding the distribution of sessions [53]. The effectiveness of this method is nearly 80%. The most important testing of the model is related to future behavior based on past training. The effectiveness of this method in the two months following calibration was about 73% of the simulated visit match in observed distributions This number exceeds traditional web mining algorithms, which actually also predict with higher accuracy, but only for the next step knowing the previous visited pages in a session [11]. The difference is that our model predicts entire distributions of visits while traditional methods predict only one step.

6. Conclusions and further work

This effort was proposed to study and apply a psychologically based theory of decision making to WUM in order to describe user's sessions. The mechanism used for such purposes was the LCA (Eq. (1)) model of decision making. It is adapted to predict the next page in a session by obtaining the sequence of visited pages. This corresponds to a stochastic simulation scheme that is handled by Monte Carlo techniques. Once the model is calibrated, it is possible to obtain the distribution of navigation trails using Monte Carlo techniques. As a sub-product of the calibration mechanism it results in the dispersion of the web user's keyword interest. With both tools it is possible to build an automatic mechanism for giving the best recommendation to web users in order to enhance the web site experience. We enforce some assumption in order to reduce the model's complexity. This appears to limit the applicability of the method, but our aim is to propose a basis for further generalization. As long as the basic rules are verified then more complex models can be built on top of it. Two main concerns should be addressed in more general web sites. First, web sites based on multimedia contents should be processed using current techniques of concept extraction. The proposed future work consists in checking the model in multimedia based web sites. The other concern relates to web sites requiring web user's higher level cognition for web browsing. A common example corresponds to the web site with learning services. Simulating navigation on this kind of web sites will require a more sophisticated AI implementation. However, many advancement in psychology could be used to explore in this direction like using more complete version of the ACT-R model [1,36,35]. However, the presented model is not simple and presents a challenge by itself. New techniques for solving such a model and data pre-processing algorithms are presented. The simulation topic is relevant for further experimentation on web site configuration. Mathematical analysis of the system is performed based on exact solutions to the unconstrained problem. Finally new perspectives are identified since human behavior modeling on the web could be applied to other human activities. The naturally high number of dimensions for the differential problem (typically 20) and non-standard border conditions results in algorithmic issues. Differential problems are commonly solved by a discrete mesh on the domain. Therefore, with such a number of dimensions, a mesh can be easily compounded of $100^{20} = 10^{40}$ cells which is computationally intractable. In spite of the intricate mathematical description, the model is based on physiological principles of decision making validated experimentally. This characteristic suffices for developing the presented model on behalf of adjusting the theoretical dynamics to the observed fact. An approach based on symbolic processing and an exact polynomial solution of the unconstrained problem were proposed to avoid a dimensional explosion. The propagator operator $U(t') = e^{t'L}$ transforms solutions on time t of the LCA equationwith the border condition, to another time $t+t'$. This operator was constructed on the basis of the LCA differential operator L and can be applied to an initial condition in order to recover the dynamics in any t. Using an exact polynomial solution of the unconstrained problem to approximate a delta on $t=\varepsilon$ and nearly ensuring the border condition, it could be propagated obtaining a solution in any t. Since solution functions are based on Hermite's polynomials, all derivation and integration could be performed exactly and symbolically. In this sense, the high dimensionality of the problem has no further influence on the symbolic problem. A maximum likelihood problem for calibrating the model's parameters is envisaged. However, distribution functions are managed symbolically in the optimization process to avoid the dimensionality problem. Again, polynomial functions are an advantage for the needed operation at the moment of finding the optimum since derivatives can be evaluated exactly. An I vector could be inferred using this technique, representing an average likelihood for text preference. However, web users should represent a distribution of different preferences. In this case the inference expands to a distribution of such a vector $P(I)$. Clustering methods help in this direction predefining a compound distribution of preference based on multivariate normal distributions. Such an approximation results in a discrete and limited number of variables. Once the parameters are fitted, the simulation of web users can be performed. Simulations are based on an exact solution for the Ornstein–Uhlenbeck stochastic equation instead of using an approximated method based on a stochastic Taylor theorem. This generates a more precise path and faster convergence for distributions. The resulting system is expected to work in a slightly modified version of the web site predicting the distribution of visits to the web site. This assumption is based on the theory that it is independent of the visited web site. This method is dependent on the correct calibration of the parameters. Simulations are in fact more sensible to the μ parameter vector representing the web user preferences in a bag-of-word space. However, as far the web site content is well known by its administrator, it is possible to correct by hand topics weight in order to adjust this parameter to known user profiles. We propose as a future work to perform simulation experiments with μ profiles provided by expert. Since the simulation successfully predicts navigation changes, changes on a web site can then be investigated for optimizing measures of web site usability. Such a possibility drastically changes the concept of an adaptive web site since it is possible to predict the impact of a change based on historical visits. Before this advancement, suggestions were performed and validated only after introducing them on the web site using trial and error. Navigational improvement is now based on an optimization method where adjustment errors can be predicted.

Acknowledgments

The authors would like to acknowledge the continuous support of the Chilean Millennium Institute of Complex Engineering Systems (ICM: P-05-004-F, CONICYT: FBO16); the Fondef IDeA project (CA12I10061) entitled, AKORI: Advanced Kernel for Ocular Research and web Intelligence.

Appendix A. Mathematical tools

A.1. Interaction matrix factorization

The partial differential equation (12) is linear when $f_i(X)=X_i$. Parameters are also assumed to fulfill [50] $\lambda_i=\lambda$ and $\kappa_i=\kappa$:

$$\frac{\partial\phi}{\partial t}=-\nabla\cdot(I\phi-\omega X\phi)+\frac{1}{2}\sigma^2\nabla^2\phi \quad \text{(A.1)}$$

$$\omega=\begin{pmatrix}\kappa & \lambda & \cdots & \lambda\\ \lambda & \kappa & \ddots & \vdots\\ \vdots & \ddots & \ddots & \lambda\\ \lambda & \cdots & \lambda & \kappa\end{pmatrix} \quad \text{(A.2)}$$

The matrix ω is a Toeplitz matrix (e.g. [29]) and must be analyzed in order to study the solution of Eq. (A.1). Further generalization of the matrix ω would consist in having equal element bands in the matrix, which is in agreement with Decision Field Theory [8]. Neuronal connections belonging to the same band are interpreted as having the same distance, whereas the ω matrix represents interactions between activation levels over distance effects. Therefore, the proposed matrix (A.2) corresponds to a circulant matrix, whose diagonal form is known (e.g. [28]). Considering that the determinant $\det(\omega-x\ 1)$ is a polynomial in x, its roots are eigenvalues. Such a polynomial factorizes easily observing that adding each remaining column $j-1$ multiplied by θ^j where $\theta^n=1$ the determinant can be factorized according to (A.3) with $\kappa+\lambda\sum_{k=1}\theta^k-x$. Note that this transformation is equivalent to a discrete Fourier transform:

$$\det(\omega-x\ 1)=\begin{pmatrix}\theta^0\left(\kappa+\lambda\sum_{k=1}^{n-1}\theta^k-x\right) & \lambda & \cdots & \lambda\\ \theta^1\left(\kappa+\lambda\sum_{k=1}^{n-1}\theta^k-x\right) & \kappa & \ddots & \vdots\\ \vdots & \vdots & \ddots & \ddots & \lambda\\ \theta^{(n-1)}\left(\kappa+\lambda\sum_{k=1}^{n-1}\theta^k-x\right) & \lambda & \cdots & \lambda & \kappa\end{pmatrix} \quad \text{(A.3)}$$

Since the determinant is linear by column an eigenvalue can be identified by factorizing $x=\kappa+\lambda\sum_{k=1}^{n-1}\theta^k=\kappa+\lambda(\theta^n-\theta)/(\theta-1)=\kappa-\lambda$ and considering $\theta^n=1$. The same occurs with column $l>1$ and adding any other column multiplied by θ^l, considering $\theta=e^{2\pi i/n}$.

Theorem 1. *The eigenvalues of ω are $x_l=\kappa+\lambda\sum_{k=1}^{n-1}\theta^{kl}$ that are reduced to two cases (Eq. (A.4)):*

$$x_l=\begin{cases}\kappa-\lambda, & l<n\\ \kappa+(n-1)\lambda, & l=n\end{cases} \quad \text{(A.4)}$$

Eigenvectors can be identified from $S^1=(\omega-(\kappa-\lambda)1)=[1]_{ij}$ and $S^2=(\omega-(\kappa+(n-1)\lambda)1)$. The first matrix S^1 spawns a degenerate space with $(n-1)$ dimension for the eigenvalue $\kappa-\lambda$. Vectors $v=[v_i]$ belonging to this space must fulfill $\sum_i v_i=0$. They can be expanded in terms of the complex root of the unity $\{e^{(2\pi i/n)kj}\}$. The second matrix S^2 has a unique null vector consisting of all components equal to one. Nevertheless, it is necessary to find an orthogonal real vector basis. The proposed complex basis (root of the unity) is orthogonal, because it corresponds to a discrete Fourier transform (see [28]). Moreover, the null vector of S^2 is orthogonal with them and corresponds to taking $j=n$. However, considering only the real component $\cos(2\pi kj/n)$ it does not reach an orthogonal relationship. Several orthogonalization procedures can be chosen (e.g. Gram–Schmidt), but using a complex rotation $e^{i\eta}$ and taking the real part shows it to be a sufficient condition for imposing orthogonality. In this case using rotation with an angle of $-\pi/4$, a real orthogonal basis can be found

$$R_{kj}=Re\left(\frac{1}{\sqrt{n}}e^{(2\pi kj/n)-i\pi/4}\right) \quad \text{(A.5)}$$

The orthogonality is recovered from discrete Fourier identities. The demonstration follows in the next lines. The Discrete Fourier Transformation (DFT) matrix is defined by

$$A=\left[\frac{1}{\sqrt{n}}e^{(2\pi i/n)kj}\right]_{kj\in\{1\cdots n\}}$$

and properties of this matrix are listed

$$A=A^T \quad \text{(A.6)}$$

$$A^{-1}=A^* \quad \text{(A.7)}$$

$$AA^*_{jk}=\sum_l^n\frac{1}{n}e^{(2\pi i/n)(j-k)l} = \begin{cases}\sum_l^n\frac{1}{n}=1, & j=k\\ \frac{1}{n}\frac{e^{2\pi i(n+1)(j-k)/n}-e^{2\pi i(j-k)/n}}{e^{2\pi i(j-k)/n}-1}=0, & j\neq k\end{cases} \quad \text{(A.8)}$$

$$A^2=\sum_l^n\frac{1}{n}e^{(2\pi i/n)(j+k)l}$$

$$= \begin{cases} \sum_l^n \frac{1}{n} = 1, & j+k \in \{n, 2n\} \\ \frac{1}{n}\frac{e^{2\pi i(n+1)(j-k)/n} - e^{2\pi i(j-k)/n}}{e^{2\pi i(j-k)/n}-1} = 0, & j+k \notin \{n, 2n\} \end{cases} \tag{A.9}$$

$$A^3 = A^* \tag{A.10}$$

$$A^4 = 1 \tag{A.11}$$

The demonstration of the inverse DFT (A.8) uses the geometric series results. In the same way (A.9) results in a matrix filled with zeros unless the band $k+j=n$ and the corner $k=j=n$ is filled with one. The cube (A.10) is obtained using the square. Using those properties and considering the real matrix $R = \frac{1}{\sqrt{2}}(e^{i\eta}A + (e^{i\eta}A)^*)$, it is possible to set η in order to recover $R^2 = 1$ obtaining real orthogonal Eigenvectors of the matrix ω:

$$\begin{aligned} R^2 &= \tfrac{1}{2}(e^{2i\eta}A^2 + 2AA^* + e^{-2i\eta}(A^2)^*) \\ &= \tfrac{1}{2}((e^{2i\eta} + e^{-2i\eta})A^2 + 2) \end{aligned} \tag{A.12}$$

If $e^{2i\eta} + e^{-2i\eta} = 0$

Then $\eta = -\pi/4$

$$\begin{aligned} R &= \frac{1}{\sqrt{2}}\left(\frac{A}{\sqrt{i}} + \sqrt{i}A^*\right) \\ &= \frac{1}{\sqrt{2n}}\left[\cos\left(\frac{2\pi kj}{n}\right) + \sin\left(\frac{2\pi kj}{n}\right)\right]_{ij \in \{1 \cdots n\}} \end{aligned} \tag{A.13}$$

Finally the following theorem holds.

Theorem 2. *The matrix ω has the following diagonal representation:*

$$\omega = RDR \tag{A.14}$$

$$R_{kj} = \frac{1}{\sqrt{2n}}\left[\cos\left(\frac{2\pi kj}{n}\right) + \sin\left(\frac{2\pi kj}{n}\right)\right]_{ij \in \{1 \cdots n\}} \tag{A.15}$$

$$D_{kj} = \begin{cases} \kappa - \lambda, & k = j < n \\ \kappa + (n-1)\lambda, & k = j = n \\ 0 & \text{otherwise} \end{cases} \tag{A.16}$$

A.2. Exact solution to the Ornstein–Uhlenbeck process

The Ornstein–Uhlenbeck process has been largely studied in statistical physics (e.g. [10]) for diffusion in fluids. A canonical set of transformations allows for reducing this equation to a simple diffusion form. Using the transformation $X \to R(X - It)$, with R as defined in (A.14), the Fokker–Planck equation is reduced to the separable equation:

$$\frac{\partial \phi}{\partial t} = \sum_{i=1}^{n} \frac{\partial}{\partial X_i}\left[\lambda_i X_i \phi + \sigma^2 \Big/ 2\frac{\partial \phi}{\partial X_i}\right] \tag{A.17}$$

$$\phi(X, t) = \prod_{i=1}^{n} \phi_i(X_i, t) \tag{A.18}$$

$$\frac{\partial \phi_i}{\partial t} = \frac{\partial}{\partial X_i}\left[\lambda_i X_i \phi_i + \sigma^2 \Big/ 2\frac{\partial \phi_i}{\partial X_i}\right] \tag{A.19}$$

Considering separating variables as in Eq. (A.18) produces n one-dimensional equations as shown in (A.19). Changing variables by $X_i \leftarrow e^{\lambda_i t}X_i$, the equation changes to (A.20). Such an expression is a 1-D diffusion equation with a time-dependent diffusion coefficient $(\sigma^2/2\lambda_i^2 e^{2\lambda_i t})$:

$$\frac{\partial \phi_i}{\partial t} = \sigma^2/2\lambda_i^2 e^{2\lambda_i t}\frac{\partial^2 \phi_i}{\partial X_i^2} \tag{A.20}$$

In such a case time can be changed in order to obtain a classical diffusion equation. Such a transformation considers a time transformation, whose derivative reproduces the time-varying coefficient from the following equation:

$$\frac{dt}{dT} = \frac{1}{\sigma^2/2\lambda_i^2 e^{2\lambda_i t}} \tag{A.21}$$

$$T(t) = \frac{\sigma^2}{2\lambda_i}(e^{2\lambda_i t} - 1) \tag{A.22}$$

Using this transformation (A.22) for time, the resulting equation is a simple diffusion in (A.23), which has the well-known heat kernel as a solution shown in (A.24)

$$\frac{\partial \phi}{\partial T} = \frac{\partial^2 \phi_i}{\partial X_i^2} \tag{A.23}$$

$$\phi_i(X_i, T) = \frac{e^{-X_i^2/4T}}{\sqrt{4\pi T}} \tag{A.24}$$

The solution of the original equation is reconstructed by applying all the performed transformations backward. The solution (A.25) obtained is a kernel for the unconstrained Ornstein–Uhlenbeck equation, on $t \to 0$ whose solution tends to be Dirac's delta δ. Eqs. (A.26) and (A.27) show the same solution in a covariant form, where $|\cdot|$ is the determinant. In this case, function over ω is obtained by applying it over its diagonal element and then transforming it back using R.

Theorem 3. *LCA equation free particle solution are*

$$\phi = \prod_{i=1}^{n} [(2\pi\sigma^2\lambda_i(e^{2\lambda_1 t} - 1))^{-1/2} e^{[R(X-It)]_i^2/2\sigma^2\lambda_i(e^{2\lambda_i t}-1)}] \tag{A.25}$$

$$\phi(X, t) = (2\pi)^{-n/2}|\Sigma(t)|^{-1} e^{(X-It)^T \Sigma(t)(X-It)} \tag{A.26}$$

$$\Sigma(t) = \frac{1}{\sigma^2(e^{2\omega t} - 1)\omega} \tag{A.27}$$

Another interesting aspect about solutions of the Fokker–Planck equation is the fact that any derivative of them is also a solution. This comes simply from the assumption that the operator L is not dependent on the variable t. In such cases, iterated derivatives on t generate Hermite polynomials multiplied by the Gaussian.

A.3. Time probability of choosing an option

$$\begin{aligned} p(t) &= \frac{\partial(1 - \overline{p}(t))}{\partial t} = -\int_\Omega \frac{\partial \phi(X, t)}{\partial t}\, dX \\ &= \int_\Omega \nabla \cdot J\, dX = \oint_{\partial\Omega} J \cdot dS \end{aligned} \tag{A.28}$$

$$J_i = \phi F_i - \sigma^2 \Big/ 2\frac{\partial \phi}{\partial X_i} \tag{A.29}$$

The Fokker–Planck equation is used in its continuity form $\partial\phi/\partial t + \nabla \cdot J = 0$ according to the flux expression (A.29). Stokes' theorem is used in the integration domain Ω of Fig. 1. The boundary of the Ω region can be decomposed into several separated sets (Eqs. (A.31) and (A.32)) according to $\partial\Omega = (\bigcup_i \Delta_i)\bigcup(\bigcup_i \Psi_i)$. The surface integral can be separated on each subset:

$$p(t) = \sum_k \oint_{\Delta_k} J \cdot dS + \sum_k \oint_{\Psi_k} J \cdot dS \tag{A.30}$$

$$\Delta_k = \{X = [X_i] | X_k = 0\} \tag{A.31}$$

$$\Psi_k = \{X = [X_i] | X_k = 1\} \tag{A.32}$$

However, in each plane Δ_i the orthogonal flux vanishes according to the reflective boundary condition (5). Only the term corresponding to the Ψ_j set from Eq. (A.30) remains. The probability density $p(i,t)$ of making the decision i in time t can be identified by restricting Eq. (A.30). Total probability decomposition (A.33) is identified with flux throughout each plane's absorbent boundary:

$$p(t)=\sum_i p(i,t) \tag{A.33}$$

In the case that the decision is i, all terms with $k\neq i$ vanish, since no flux flows over Ψ_k, in which case this probability is given by the total flux over the surface Ψ_i.

Theorem 4.

$$p(i,t)=-\frac{\sigma^2}{2}\int_0^1\cdots\int_0^1 \left.\frac{\partial\phi}{\partial X_i}\right|_{X_i=1}\prod_{k\neq i} dX_k \tag{A.34}$$

The expression (A.34) is derived using the border condition $\phi=0$ on Ψ.

A.4. The reflective operator

Every linear operator that commutes with the operator L in the Fokker–Planck equation (13) transforms a solution into another. This fact is used by well-known methods in partial differential equations to generate solutions that accomplish a border condition. If the probability density $\phi_a(x,t)$ of a 1-D Brownian motion that starts on $t=0$ at $a>0$ is given, then the density of the system constrained by a reflecting barrier on $x=0$ is given by

$$\phi_a^R=\frac{1}{c}[\phi_a(x,t)+\phi_{-a}(-x,t)] \tag{A.35}$$

The border condition for reflection on $x=0$ is $(d\phi_a^R/dx)|_{x=0}=0$, which means that no probability flux crosses the barrier $x=0$. In this sense, $\phi_a{}^R$ satisfies both the border condition and the diffusion equation. Nevertheless, normalization is satisfied since both terms in (A.35) are normalized and then need to be divided by $c=\int_{x>0}[\phi_a(x,t)+\phi_{-a}(-x,t)]\,dx$. Such a term is independent of t since dc/dt by the diffusion equation integrates a derivative that cancels on borders $\{0,\infty\}$. This result can be generalized to several dimensions. The important point in the previous derivation was the cancelation of derivatives on the barrier $x=0$. Furthermore, a transformation $U_1(x)=-x$ was applied over the parameter a and variable x in order to change the sign of the derivative for canceling on the barrier. The operator U_1 is linear since it belongs to the reflection group. In several dimensions, the whole reflection group over reflective planes needs to be considered for enabling the perpendicular gradient cancelation on planes $\Delta_i=\{X\in\mathbb{R}^n \mid X_i=0\}$. The reflection group is constructed by using representation of linear transformation $Q_i(X)$, where $[Q_i(X)]_i=-X_i$ and $[Q_i(X)]_{j\neq i}=X_j$. The closure is then built by composition. The complete group consists of the set $\Re=\{Q_{i_1}\ldots Q_{i_g} \mid i_a\neq i_b,\ a\neq b,\ g=1,\ldots,n\}$. Some properties are well known, $\forall\ A,B\in\Re$, then $A^2=1$ and $AB=BA$.

Theorem 5. *If $\phi(X,t)$ satisfies the diffusion equation on $\mathbb{R}^n$, then a solution $\phi^R(X,t)$ that fulfills reflecting conditions on all planes Δ_i (or $(\partial U_n[\phi(X,t)]/\partial X_i)|_{X_i=0}\ \forall i$) can be constructed as*

$$U_n[\phi(X,t)]=\sum_{Q\in\Re}\phi(QX,t) \tag{A.36}$$

The demonstration of such a construction is by induction on n, since it holds for $n=1$. And using the reduction $QX|_{X_i=0}=Q'X|_{X_i=0}$, where Q' is a product of all components of Q except Q_i. Moreover, if when restricted to Δ_i some terms of this sum (A.36) reduce the dimension by one, then induction can be applied. For each element Q that contains a Q_i when restricted to Δ_i, it is equivalent to Q' so $Q=Q'Q_i$. In such cases,

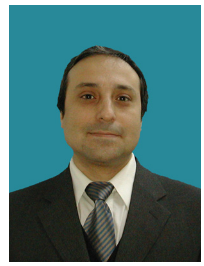
$$\left.\frac{\partial\phi(QX,t)}{\partial X_i}\right|_{X_i=0}=-\left.\frac{\partial\phi(Q'X,t)}{\partial X_i}\right|_{X_i=0}$$

and both terms cancel in (A.36) because $Q'\in\Re$. On the other hand, if a term Q' does not contain Q_i then there exists a term $Q=Q'Q_i\in\Re$ that will cancel. Furthermore, the function $U_n[\phi(X,t)]$ satisfies reflection conditions on Δ as $(\partial U_n[\phi^R(X,t)]/\partial X_i)|_{X\in\Delta_i}=0$, and it is a solution of the diffusion equation due to invariance to transformation $X\to QX$, because $Q^2=1$. Operator U_n is linear as long as it is a linear combination of linear reflection operators.

References

[1] J. Anderson, D. Bothell, M. Byrne, S. Douglass, C. Lebiere, Y. Qin, An integrated theory of the mind, Psychol. Rev. 111 (2004) 1036–1060.
[2] R.J. Audley, A.R. Pike, Some alternative stochastic models of choice, Br. J. Math. Stat. Psychol. 18 (1965) 207–225.
[3] A. Beskos, O. Papaspiliopoulos, G.O. Roberts, Retrospective exact paths with applications, Bernoulli 12 (2006) 1077–1098.
[4] A. Beskos, G.O. Roberts, Exact simulation of diffusions, Ann. Appl. Probab. (2005) 2422–2444.
[5] K. Beyer, J. Goldstein, R. Ramakristan, U. Shaft, When is nearest neighbor meaningful, Database Theory 15 (1999) 217–235.
[6] A. Blum, H. Chan, M.R. Rwebangira, A random-surfer web-graph model, in: The Eighth Workshop on Algorithm Engineering and Experiments and the Third Workshop on Analytic Algorithmics and Combinatorics, SIAM, 2006, pp. 238–246.
[7] M. Bronstein, Symbolic Integration I: Transcendental functions, Algorithms and Computation in Mathematics, Springer, 2005.
[8] J.R. Busemeyer, J.T. Townsend, Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment, Psychol. Rev. 100 (1993) 432–459.
[9] J.R. Busemeyer, Z. Wang, J.T. Townsend, A quantum dynamics of human decision making, J. Math. Psychol. 50 (2006) 220–241.
[10] S. Chandrasekhar, Stochastic problems in physics and astronomy, Rev. Mod. Phys. 15 (1943) 1–89.
[11] M. Deshpande, G. Karypis, Selective Markov models for predicting web page accesses, ACM Trans. Internet Technol. 4 (2004) 163–184.
[12] P.L. Emerson, Simple reaction time with Markovian evolution of Gaussian discriminal processes, Psychometrika 35 (1970) 99–109.
[13] D.T. Gillespie, Exact numerical simulation of the Ornstein–Uhlenbeck process and its integral, Phys. Rev. E 54 (1996) 2089–2091.
[14] T.D. Hanks, J. Ditterich, M.N. Shadlen, Microstimulation of macaque area lip affects decision-making in a motion discrimination task, Nat. Neurosci. 9 (2006) 682–689.
[15] H.R. Heekeren, S. Marrett, L.G. Ungerleider, The neural systems that mediate human perceptual decision making, Nat. Rev. Neurosci. 9 (2008) 467–479.
[16] B.A. Huberman, P.L.T. Pirolli, J.E. Pitkow, R.M. Lukose, Strong regularities in world wide web surfing, Science 280 (1998) 95–97.
[17] R. Kiani, M.N. Shadlen, Representation of confidence associated with a decision by neurons in the parietal cortex, Science 324 (2008) 759–764.
[18] J.N. Kim, M.N. Shadlen, Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque, Nat. Neurosci. 2 (1999) 176–185.
[19] M. Kitajima, M.H. Blackmon, P.G. Polson, Cognitive architecture for website design and usability evaluation: comprehension and information scent in performing by exploration, in: Proceedings of the Human–Computer Interaction International Conference, 2005, pp. 343–373.
[20] M. Kitajima, P. Polson, M. Blackmon, CoLiDeS and SNIF-ACT: complementary models for searching and sensemaking on the web, in: Human–Computer Interaction Consortium (HCIC) Winter Workshop, 2007, p. 1.
[21] D. Laberge, A recruitment theory of simple behavior, Psychometrika 27 (1979) 375–396.
[22] D.R.J. Laming, Information Theory of Choice Reaction Time, Wiley, 1968.
[23] C. Lemieux, Monte Carlo and Quasi-Monte Carlo Sampling, Springer Series in Statistics, Springer, 2009.
[24] P. Loyola, P.E. Román, J.D. Velásquez, Clustering-based learning approach for ant colony optimization model to simulate web user behavior, in: 2011 IEEE/WIC/ACM International Conference, vol. 1, France, pp. 457–464. doi: http://dx.doi.org/10.1109/WI-IAT.2011.116.
[25] P. Loyola, P.E. Román, J.D. Velásquez, Predicting web user behavior using learning-based ant colony optimization, J. Eng. Appl. Artif. Intell. 25 (2011) 889–897.
[26] R. Luce, P. Suppes, Preference, utility and subjective probability, in: R.D. Luce, R.R. Bush, E. Galanterá (Eds.), Handbook of Mathematical Psychology, vol. III, Wiley, 1965, pp. 249–410.
[27] C.D. Manning, H. Schutze, Foundation of Statistical Natural Language Processing, The MIT Press, 1999.

[28] M.L. Mehta, Matrix Theory: Selected Topics and Useful Results, Les edition de physique, 1988.
[29] C. Meyer, Matrix Analysis, SIAM, 2001.
[30] C.S. Miller, R.W. Remington, Modeling information navigation: implications for information architecture, Hum. Comput. Interact. 19 (2004) 225–271.
[31] B. Mobasher, H. Dai, T. Luo, M. Nakagawa, Discovery and evaluation of aggregate usage profiles for web personalization, Data Min. Knowl. Discov. 6 (2002) 61–82.
[32] O. Nasraoui, M. Soliman, E. Saka, A. Badia, R. Germain, A web usage mining framework for mining evolving user profiles in dynamic web sites, IEEE Trans. Knowl. Data Eng. 20 (2008) 202–215.
[33] W. Newey, D. McFadden, Large sample estimation and hypothesis testing. Handbook of Econometrics, vol. 4, North Holland, Amsterdam, 1994, pp. 2111–2245.
[34] B.K. Oksendal, Stochastic Differential Equations: An Introduction with Applications, 5 ed., Springer, 2002.
[35] P.L.T. Pirolli, Power of 10: modeling complex information-seeking systems at multiple scales, Computer 42 (2009) 33–40.
[36] P.L.T. Pirolli, W.T. Fu, SNIF-ACT: a model of information foraging on the world wide web, in: The Ninth International Conference on User Modeling, 2003, pp. 45–54.
[37] R. Ratcliff, A theory of memory retrieval, Psychol. Rev. 85 (1978) 59–108.
[38] S.I. Resnick, Adventures in Stochastic Processes, Birkhauser Verlag, Basel, Switzerland, 1992.
[39] J.D. Roitman, M.N. Shadlen, Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task, J. Neurosci. 22 (2002) 9475–9489.
[40] P.E. Román, Web user behavior analysis (Ph.D. thesis), University of Chile, 2011.
[41] P.E. Román, G. L'Huillier, J.D. Velásquez, Web usage mining, in: J.D. Velásquez, L. Jain (Eds.), Advanced Techniques in Web Intelligence, Springer, 2010, pp. 143–165.
[42] P.E. Román, J.D. Velásquez, Analysis of the web user behavior with a psychologically-based diffusion model, in: AAAI 2009 Fall Symposium on Biologically Inspired Cognitive Architectures, Arlington, Washington DC, USA, Technical Paper of the AAAI.
[43] P.E. Román, J.D. Velásquez, Stochastic simulation of web users, in: Proceedings of the 2010 IEEE/WIC/ACM International Conference, IEEE Press, Toronto, Canada, 2010, pp. 30–33.
[44] P.E. Román, J.D. Velásquez, V. Palade, L.C. Jain, New trends in web user behaviour analysis, in: J.D. Velásquez, V. Palade, L.C. Jain (Eds.), Advanced Techniques in Web Intelligence-2, Studies in Computational Intelligence, vol. 452, Springer, Berlin Heidelberg, 2013, pp. 1–10.
[45] T. Sato, A. Murthy, K.G. Thompson, J.D. Schall, Search efficiency but not response interference affects visual selection in frontal eye field, Neuron 30 (2001) 583–591.
[46] J.D. Schall, Frontal eye fields, in: Encyclopedia of Neuroscience, vol. 4, Elsevier, 2009, pp. 367–374.
[47] M. Stone, Models for choice reaction time, Psychometrika 25 (1960) 251–260.
[48] L. Tauscher, S. Greenberg, Revisitation patterns in world wide web navigation, in: Proceedings of the Conference on Human Factors in Computing Systems, Atlanta, USA, pp. 22–27.
[49] K. Train, Discrete Choice Methods with Simulation, Cambridge University Press, 2009.
[50] M. Usher, J.L. McClelland, The time course of perceptual choice: the leaky, competing accumulator model, Psychol. Rev. 2 (2001) 550–592.
[51] J.D. Velásquez, V. Palade, Testing online navigation recommendations in a web site, in: B. Gabrys, R.J. Howlett, L.C. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Lecture Notes in Computer Science, vol. 4253, Springer, Berlin Heidelberg, 2006, pp. 487–496.
[52] J.D. Velásquez, V. Palade, Building a knowledge base for implementing a web-based computerized recommendation system, Int. J. Artif. Intell. Tools 16 (2007) 793–828.
[53] J.D. Velásquez, V. Palade, A knowledge base for the maintenance of knowledge extracted from web data, J. Knowl. Based Syst. 20 (2007) 238–248.
[54] J.D. Velásquez, V. Palade, Adaptive Web Sites: A Knowledge Extraction from Web Data Approach, IOS Press, Amsterdam, NL, 2008.
[55] J.D. Velásquez, V. Palade, L.C. Jain (Eds.), Advanced Techniques in Web Intelligence – 2, number v. 1 in Studies in Computational Intelligence, 1 ed., Springer, 2013.



Pablo E. Román has been an Adjunct Instructor Lecturer since 2004 and researcher with the Department of Industrial Engineering, School of Engineering and Science, University of Chile. He received the BS degree in Physics in 1992, the B.S. degree and Ing. in Computer Science in 1996 and 1999 respectively, the Ph.D. degree in Engineering Systems from the University of Chile in March 2011. He is a reviewer of journals and conferences and co-author of the series "Advanced Techniques in Web Intelligence" parts 1 and 2 (Springer, 2010, 2012). Currently, he is Postdoctoral Researcher at the Center of Mathematical Modeling of the University of Chile. He is a IEEE member. His current research interests are Scientific Computing, Astroinformatics, Neurocomputing, simulations of human behavior, and Web Usage Mining.

Juan D. Velásquez is an Associate Professor with the Department of Industrial Engineering, School of Engineering and Science, University of Chile. He received his B.E. in Electrical Engineering and B.E. in Computer Science in 1995, P.E. in Electrical Engineering and P.E. in Computer Engineering in 1996, Masters in Computer Science, Industrial Engineering and Informatics Law in 2001, 2002 and 2010, respectively, from the University of Chile. He got his Ph.D. from the University of Tokyo, Japan in 2005. He has been a Postdoctoral Fellow with the Computing Laboratory, University of Oxford, UK in 2005 and a Visiting Professor with the Center for Collaborative Research, University of Tokyo, Japan in 2006 and 2007. He is a reviewer of journals and conferences and he has written approximately 100 papers for conferences and journals, five book chapters and he is the co-author of the book "Adaptive Web Sites" (IOS Press, 2008), editor and co-author of the book series "Advanced Techniques in Web Intelligence part 1 and part 2" (Springer, 2010, 2012). His research interests are web intelligence, knowledge discovery from web data and web user behavior analysis.