Expert Systems with Applications 41 (2014) 6327–6345

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Behavior monitoring under uncertainty using Bayesian surprise and optimal action selection

Luis Avila, Ernesto Martínez *

INGAR (CONICET-UTN), Avellaneda 3657, Santa Fe S3002 GJC, Argentina

ARTICLE INFO

Keywords:
Artificial pancreas
Bayesian surprise
Behavior monitoring
Kullback–Leibler divergence
Optimal action selection

ABSTRACT

The increasing trend towards delegating tasks to autonomous artificial agents in safety–critical socio-technical systems makes monitoring an action selection policy of paramount importance. Agent behavior monitoring may profit from a stochastic specification of an optimal policy under uncertainty. A probabilistic monitoring approach is proposed to assess if an agent behavior (or policy) respects its specification. The desired policy is modeled by a prior distribution for state transitions in an optimally-controlled stochastic process. Bayesian surprise is defined as the Kullback–Leibler divergence between the state transition distribution for the observed behavior and the distribution for optimal action selection. To provide a sensitive on-line estimation of Bayesian surprise with small samples twin Gaussian processes are used. Timely detection of a deviant behavior or anomaly in an artificial pancreas highlights the sensitivity of Bayesian surprise to a meaningful discrepancy regarding the stochastic optimal policy when there exist excessive glycemic variability, sensor errors, controller ill-tuning and infusion pump malfunctioning. To reject outliers and leave out redundant information, on-line sparsification of data streams is proposed.

1. Introduction

Safety–critical systems integrate software, hardware and humans in an increasing number of applications that are prone to failures, errors and malfunctioning which could result in loss of life, significant property damage, or damage to the environment. Moreover, the increasing trend towards delegating tasks to autonomous artificial agents in safety–critical socio-technical systems makes monitoring an action selection policy of paramount importance. As an example, consider the case of collision avoidance in driving systems (Broggi, Medici, Zani, Coati, & Panciroli, 2012) where the monitoring task involves a number of autonomous vehicles interacting with each other in a high-speed highway. Any monitoring system aimed to warn or prevent collisions and dangerous circumstances must contemplate the expected behavior of nearby cars to detect quickly a collision scenario. However, monitoring tasks are often formulated around the idea of isolated agents with perfect rationality (Thimbleby, 2009). For on-line traffic monitoring, the key aspect is to characterize the uncertain environment the autonomous car is in and its desired optimal behavior. Similarly, there is an increased requirement for condition monitoring of nuclear power plants to ensure they are still able to operate safely, yet efficiently (West, McArthur, & Towle, 2012). Decision support to detect anomalies is limited by the availability of expert knowledge and the variability of the plant conditions. Proper control and on-line monitoring of the interaction between the operators and the plant would be helpful to prevent catastrophic accidents (Salge & Milling, 2006). In a different field, researchers have evaluated different solutions to automate the task of gazing at a monitor to find suspicious behaviors in video surveillance (Fernández-Caballero, Castillo, & Rodríguez-Sánchez, 2012). Detecting dangerous objects and intruders is essential for safety in crowded environments, but monitoring human behaviors and reporting about anomalies is a complex task for any computing system.

Current automated systems function well in environments they are designed for, that is, around their nominal operating conditions or expected scenarios. They also perform well in environments with "predictable" uncertainties as treated, for example, in the advanced adaptive and robust control frameworks. Nevertheless, control systems of today require substantial human intervention when faced with novel and unanticipated situations, i.e. situations that have not been considered at the design stage. Such situations can arise from discrete changes in the environment, extreme disturbances, structural changes in the system (for example, as a result of damage), and the like. More specifically, biological control systems such as the artificial pancreas (AP) must face significant

* Corresponding author. Tel.: +54 (342) 4534451; fax: +54 (342) 4553439.
E-mail address: ecmarti@santafe-conicet.gov.ar (E. Martínez).

http://dx.doi.org/10.1016/j.eswa.2014.04.031
0957-4174/

Nomenclature

Symbols for the glycemic model

λ	drift parameter
σ	variability parameter
BG	blood glucose level
G_{in}	systemic appearance of glucose via glucose absorption from the gut
$NHGB$	net hepatic glucose balance
G_{out}	overall rate of peripheral and insulin dependent glucose utilization
G_{ren}	excretion of glucose
V_G	volume of distribution of glucose
Sh	hepatic sensitivity
Sp	insulin sensitivity
I	insulin infusion level
IG	interstitial glucose
τ	sensor time-lag parameter
ξ	sensor calibration parameter
k_C	PID proportional gain
τ_I	PID integral time
τ_D	PID derivative time

Symbols for the control algorithm

m	mean function
cov	covariance function
$p(\bullet\vert\bullet,\bullet)$	controlled transition probability
$\hat{\mathbf{x}}$	state estimation
Δx	state change
u	control action
$\pi(\bullet)$	policy
$V(\bullet)$	state function value
$Q(\bullet,\bullet)$	state-action function value
$\ell(\bullet,\bullet)$	immediate cost
$h(\bullet\vert\bullet)$	passive dynamics
γ	discount factor
r	reward function
GP	Gaussian process
ϑ	cardinality
ρ	exploration parameter
β	exploitation parameter
$\mathfrak{x}$	state space
U	action space

Symbols for monitoring task

$P(\bullet)$	prior probability
$P(\bullet\vert\bullet)$	posterior probability
$KL(\bullet\Vert\bullet)$	Kullback–Leibler distance
P_{KL}	pointwise Bayesian surprise
T_{KL}	robust Bayesian surprise
$\mathrm{k}(\bullet,\bullet)$	kernel
δ	stop threshold
η	threshold for the level of sparsity
$N_{\max}$	size of the training set used to model the Gaussian process
$\mathfrak{D}$	dictionary

levels of variability. When an action is executed by an agent, the perceived result of the action depends on the environmental response, including other agents, noisy measurements, hidden states and the quality of the sensory data. In most cases, the agent has only an approximate knowledge of these effects, but it must nevertheless choose a nearly-optimal course of action to accomplish the desired control task (Sanger, 2011). Under uncertainty, a probabilistic characterization of the desired behavior is needed to assess if a given agent behavior respects its specification. Such a specification is an essential element of using Bayesian inference to detect deviations from an optimal control policy.

Almost all of the existing literature about system monitoring, is concerned with the task to make certain controlled variables track given set-points or set-point trajectories, while assuring closed-loop stability. However, the purpose of autonomy (and that of automation as a whole), is not primarily to keep the controlled variables at their set-points as well as possible or to nicely track dynamic set-point changes. For example, a feasible controller for glycemic regulation based on model predictive control has been designed to control to a zone instead of a set-point, which may prevent unnecessary and dangerous overcorrection (Grosman, Dassau, Zisser, Jovanoviĉ, & Doyle, 2010). An important issue is that the agent decision-making policy is mainly focused on the net return which should be maximized in the presence of disturbances and different sources of variability, while exploiting the available noisy and scarce measurements. Thus, behavior monitoring under uncertainty has to be built upon a stochastic process specification of the desired optimal policy.

The novelty and relevance of information contained in new data, can be measured by the effect such data has on the observer (monitor) (Hasanbelliu, Kampa, Principe, & Cobb, 2012). Fundamentally, this effect is to transform the observer's prior beliefs into posterior beliefs, according to the Bayes theorem. The amount of information can be measured in a natural way by the Kullback–Leibler (KL) distance -also called relative entropy- between the prior and posterior distributions in the observer, regarding the available space of hypotheses about the state of a controlled system. This facet of information, termed "surprise," is important in behavior monitoring where beliefs change over time, in particular when malfunctioning causes a deviant behavior. Surprise is a subjective information measure that quantifies how much information a new observation contains, in relation to the current knowledge of the system being monitored (Baldi & Itti, 2010; Itti & Baldi, 2005b). Surprise can exist only in the presence of uncertainty, and it is related to beliefs about the dynamics of state transitions, where the same data convey different amount of surprise to different observers or to the same observer at different times. To quantify the surprise factor of an observation, in this work the novelty of information in a data stream regarding deviations from the specified behavior is measured using twin Gaussian processes (Bo & Sminchisescu, 2010).

Behavior specification under uncertainty is formalized here as a controlled stochastic process that makes the agent policy as close as possible to the desired one by describing both the policy and the state transition dynamics in probabilistic terms. To this aim, *optimal choice of actions under uncertainty* is a fundamental problem to be addressed in order to characterize the desired behavior of an intelligent agent. The abstract setting for the latter can be framed as an agent choosing actions over time, an uncertain dynamical system whose state is affected by those actions, and a performance criterion that the agent seeks to optimize (Todorov, 2009). The agent has the power to reshape the system dynamics in any way it wishes. However, it pays a price for too much reshaping (Dvijotham & Todorov, 2012). The key question for on-line behavior monitoring is how the "distance" from optimal reshaping can be measured using small samples from realizations of a