Engineering Applications of Artificial Intelligence 25 (2012) 1515–1522

Contents lists available at SciVerse ScienceDirect

Engineering Applications of Artificial Intelligence

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

Iterative learning control based tools to learn from human error

Philippe Polet[a,b,c,*], Frédéric Vanderhaegen[a,b,c], Stéphane Zieba[d]

[a] Univ Lille Nord de France, F-59000 Lille, France
[b] UVHC, LAMIH, F-59313 Valenciennes, France
[c] CNRS, FRE 3304, F-59313 Valenciennes, France
[d] Laboratory for Cognitive Systems Science, Department of Risk Engineering, University of Tsukuba, Tsukuba, Japan

ARTICLE INFO

Article history:
Received 3 June 2011
Received in revised form
3 January 2012
Accepted 11 January 2012
Available online 28 February 2012

Keywords:
Human factors
Iterative learning control
Reinforcement
Reliability
Utility function

ABSTRACT

This paper proposes a new alternative to identify and predict intentional human errors based on benefits, costs and deficits (BCD) associated to particular human deviations. It is based on an iterative learning system. Two approaches are proposed. These approaches consist in predicting barrier removal, i.e., non-respect of rules, achieved by human operators and in using the developed iterative learning system to learn from barrier removal behaviours. The first approach reinforces the parameters of a utility function associated to the respect of this rule. This reinforcement affects directly the output of the predictive tool. The second approach reinforces the knowledge of the learning tool stored into its database. Data from an experimental study related to driving situation in car simulator have been used for both tools in order to predict the behaviour of drivers. The two predictive tools make predictions from subjective data coming from drivers. These subjective data concern the subjective evaluation of BCD related to the respect of the right priority rule.

1. Introduction

Human reliability is be defined as the capacity of the human operators to achieve their required tasks in predefined conditions and to not achieve additional tasks that may damage the system safety (Swain and Guttman, 1983). Human error is the complementary concept: it is the capacity of human operators to not achieve their required tasks in predefined conditions or to achieve additional tasks that may damage the system safety. Human reliability or human error relate then to several principles: the human task analysis and modelling, the erroneous task identification, the error evaluation.

A lot of human error analysis methods exist, but most of them have several problems to be solved (Vanderhaegen, 2003):

- The predefined conditions concerning the hypotheses to identify and assess human errors are not well-defined. For instance, the related capacity of human operators to achieve tasks can lead to a period of time or an instantaneous time, can take into account the human experience, can include or not the human recovery process.
- The measure of this capacity is usually assimilated as a probability of occurrence of a human error. Nevertheless, the conditions with which the probability was assessed are usually not described. The units of the probability are not specified and it seems difficult to compare probabilities that were obtained with different units. For instance, the ratio of the number of human error occurrence upon the number of solicitations of the same tasks cannot be compared with a ratio of the number of human error occurrence upon time unit. When the human error methods consist in assessing the risks associated to the human errors, they include a measure of the human error consequences. Here again, the units taken for the consequence assessment are not always clearly defined.
- During the task analysis and the tasks modelling processes, these methods do not take into account all the dependencies between tasks such as functional dependencies (Vanderhaegen et al., 1994), time dependencies (Vanderhaegen, 1999) or causal dependencies (Vanderhaegen, 2004).
- Results obtained with the human reliability analysis methods are often not homogeneous (Swain, 1990; Kirwan, 1997), are limited to unintentional human errors and are off-line methods (Vanderhaegen, 2003).
- Moreover typology of errors taken into account by the majority of methods deals with lapses, mistakes and faults. These kinds of errors are unintentional errors; violations are rarely taken into account.
- Comparison between a priori risk analysis and a posteriori analysis reveals some differences. These differences may be explain by violations commissions and a gap between the conditions of use of the system at the design stage and the real conditions of use constrained by the context (evolution of productivity demand, variability of crew, etc.) (Amalberti, 2001; Rasmussen, 1997; Fadier and la Garza, 2007).

* Corresponding author.
E-mail addresses: philippe.polet@univ-valenciennes.fr, ppolet@univ-valenciennes.fr (P. Polet), vanderhaegen@univ-valenciennes.fr (F. Vanderhaegen), zieba@css.risk.tsukuba.ac.jp (S. Zieba).

0952-1976/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.engappai.2012.01.010

In the literature dedicated to car-driving violations, the main approach is the statistical classification: different studies try to find the main characteristics of a driving leading to commission of violation. These characteristics are the gender, the age, mean age and psychological aspects such as sensation seeking, etc. (Lucidi et al., 2010). It is a classification way and cannot be used for on-line prediction.

This paper consists in proposing an on-line approach to predict intentional human error without taking into account any probability. This approach is based on the iterative learning control concept. The iterative learning control concept was initially used to learn from errors when achieving automated repetitive tasks (Lee et al., 2000; Xu and Yan, 2004; Xu et al., 2004; Chien and Yao, 2004; Norrlöf and Gunnarsson, 2005). It is adapted to develop a prediction system and to test the feasibility to predict particular intentional human errors called barrier related violations or barrier removals by taking into account the consequences of these human erroneous behaviours.

A barrier is a human or technical system that aims at protecting the human–machine system from the occurrence or the consequences of undesirable events. It can be material such as a wall, or immaterial such as a procedure. A barrier is usually attached to a function. For instance, safety barriers are designed to protect the system from unsafe events and reliability barriers are designed to protect the system from unreliable events. Sometimes, human operators on field decide to not respect these barriers: this kind of erroneous behaviour is called barrier removal (Polet et al., 2009).

The barrier removal concept was already studied and observed for different application domains:

- Barrier removal during the use of production system such as industrial rotary press (Polet et al., 2002).
- Barrier removal during the control of transport system such as car driving (Chaali-Djelassi and Vanderhaegen, 2006) or train control (Vanderhaegen et al., 2002; Polet and Vanderhaegen, 2007).
- Barrier removal of biomechanical applications such as human behaviour in crash context (Robache et al., 2006; Pacaux-Lemoine and Vanderhaegen, 2007).

These studies propose that the human decision to respect or not a barrier depends on 3 attributes: the benefits, the costs and the potential deficits associated to the non-respect of the barrier (Polet et al., 2002).

2. Barriers and BCD modelling

With this BCD model, human operator decision-making becomes a process of balancing the advantages and disadvantages of removing a barrier (Polet et al., 2009). In order to describe this balancing process more precisely, it is necessary to adopt a multi-criteria viewpoint. Clearly, the behaviour of human operators has consequences on safety, but also on other criteria, such as productivity, quality and workload, to name a few. These criteria depend on the nature of the activity.

The benefit and cost factors take into account the consequences of barrier-removal if it happens normally, with no deviate events, while the deficit factor evaluates the consequences of the barrier-removal if deviate events occur. The deficit factor is related to the potential risk associated with barrier-removal. Determining benefit, cost and deficit requires evaluating the consequences of operator behaviour. When evaluating the consequences of a given behaviour, either a designer-specified behaviour or a barrier removal, two cases are possible. The first considers the success of the behaviour, and the result of the evaluation is denoted CS. The second considers the failure of the behaviour, and the result is denoted CF. These evaluations are done by human operators and constitute a sort of risk evaluation procedure. Since the evaluation takes several criteria into account, CS and CF are vectors. The analysis of a human operator's behaviour can integrate or ignore the possible interference of other behaviours by the same human operator or by other human operators. To calculate the benefit of a barrier removing, $B(\mathrm{BR})$, cost $C(\mathrm{BR})$ and deficit $D(\mathrm{BR})$, two behaviours must be evaluated: the designer-specified behaviour-evaluated (i.e., a non-barrier removing, NBR) using CS(NBR) and CF(NBR) – and the barrier-removal behaviour – evaluated using CS(BR) and CF(BR).

Benefits, costs and deficits are positive values. Given CS(NBR), CF(NBR), CS(BR) and CF(BR), the vectors B(BR), C(BR) and D(BR) can be calculated as follows (cf. Eq. (1)):

$$\forall i, \tag{1}$$

$$\begin{aligned}
&\text{if } (\mathrm{CS(BR)}_i - \mathrm{CS(NBR)}_i) > 0\\
&\text{then } B(\mathrm{BR})_i := \mathrm{CS(BR)}_i - \mathrm{CS(NBR)}_i, \text{ and } C(\mathrm{BR})_i := 0\\
&\text{else } B(\mathrm{BR})_i := 0, \text{ and } C(\mathrm{BR})_i := \mathrm{CS(NBR)}_i - \mathrm{CS(BR)}_i
\end{aligned}$$

$$\begin{aligned}
&\text{if } (\mathrm{CF(BR)}_i - \mathrm{CF(NBR)}_i) < 0\\
&\text{then } D(\mathrm{BR})_i := \mathrm{CF(NBR)}_i - \mathrm{CF(BR)}_i\\
&\text{else } D(\mathrm{BR})_i := 0
\end{aligned}$$

where i is the criterion considered.

The possibility to predict the barrier removing by observing objective data has been already studied (Vanderhaegen et al., 2009). The context of this study was the car driving. The predictive tool was based on iterative learning control approach.

3. Barrier removing prediction based on iterative learning control approach

The reinforcement is a major function of iterative learning control (Bucak and Zohdy, 2001; Wang and Usher, 2005; Duan et al., 2007). It consists in managing the positive and negative error of learning in order to achieve predefined control goals. This paper proposes two approaches using two different process of reinforcement. For the first one, the parameters for prediction are reinforced. For the second one, knowledge used by the tool is reinforced.

The two approaches aim at predicting intentional human error without taking into account any probability. These approaches are based on the iterative learning control concept.

The principle of the iterative learning control approach is decomposed in 3 steps:

1. to calculate the output u_i^* regarding the input e_i by a predictive tool,
2. to evaluate the predictive tool output and
3. to correct the parameters of the predictive tool.

The input vector (e_i) at an iteration I is composed by the benefits, costs and deficits:

$$e_i = (b_1, c_1, d_1 \ldots b_n, c_n, d_n)$$

The output (u_i^*) is a binary value corresponding to a barrier removal (BR) or a non-barrier removal (NBR) (Fig. 1).

In the study presented in this paper, the prediction concerns the behaviour of a driver in a particular situation where a rule (the rule is then considered as a barrier) has to be respected. The predicted behaviour is then compared to the real one. Depending