Artificial Intelligence in Medicine 48 (2010) 1–19

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

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

Adopting model checking techniques for clinical guidelines verification

Alessio Bottrighi [a], Laura Giordano [a], Gianpaolo Molino [b], Stefania Montani [a], Paolo Terenziani [a,*], Mauro Torchio [b]

[a] Dipartimento di Informatica, Università del Piemonte Orientale "Amedeo Avogadro", Viale Teresa Michel 11, 15121 Alessandria, Italy
[b] Azienda Ospedaliera San Giovanni Battista, corso Bramante 88, 10126 Torino, Italy

ARTICLE INFO

Article history:
Received 11 October 2007
Received in revised form 14 September 2009
Accepted 14 September 2009

Keywords:
Clinical guidelines
Model checking
Verification

ABSTRACT

Objectives: Clinical guidelines (GLs) are assuming a major role in the medical area, in order to grant the quality of the medical assistance and to optimize medical treatments within healthcare organizations. The verification of properties of the GL (e.g., the verification of GL correctness with respect to several criteria) is a demanding task, which may be enhanced through the adoption of advanced Artificial Intelligence techniques. In this paper, we propose a general and flexible approach to address such a task.
Methods and materials: Our approach to GL verification is based on the integration of a computerized GL management system with a model-checker. We propose a general methodology, and we instantiate it by loosely coupling GLARE, our system for acquiring, representing and executing GLs, with the model-checker SPIN.
Results: We have carried out an in-depth analysis of the types of properties that can be effectively verified using our approach, and we have completed an overview of the usefulness of the verification task at the different stages of the GL life-cycle. In particular, experimentation on a GL for ischemic stroke has shown that the automatic verification of properties in the model checking approach is able to discover inconsistencies in the GL that cannot be detected in advance by hand.
Conclusion: Our approach thus represents a further step in the direction of general and flexible automated GL verification, which also meets usability requirements.

1. Introduction

Clinical guidelines (GLs) can be defined as a means for specifying the "best" clinical procedures and for standardizing them. The adoption of GLs, by supporting physicians in their decision making and diagnosing activities, may provide crucial advantages, both in individual-based health care, and in the overall service offered by a health care organization. In particular, it has been shown [1] that GLs can improve the quality of patient care, reduce variations in quality of care, and reduce costs.

These observations justify the increasing number of GLs which have been defined in the last decade, covering a large spectrum of diseases and medical procedures. However, the effort in defining and disseminating GLs has not always been coupled by a parallel effort in guaranteeing their "quality" [2]: despite the fact that GLs are issued by recognized experts' committees, they might be ambiguous or incomplete [3], or even inconsistent.

The need for GL quality *verification* is thus clearly emerging. As we will show in this paper, computer-based approaches can provide crucial advantages in this context. The research community in Artificial Intelligence (AI) in medicine and in medical decision making, which is very active in the definition of computerized systems and projects for managing GLs (see e.g. the systems Asbru [4], EON [5], GEM [6], GLARE [7–8], GLIF [9], GUIDE [10], PROforma [11], and the collections [12–14]), has recently started to consider this issue.

Nevertheless, the verification capabilities available in the conventional computerized GL management systems in the literature are usually rather limited and only recently this limitation has led to the development of proposals for guideline automatic verification. Let us first analyse the limitations of conventional computerized GL management systems. In many cases, such systems do associate only very specific and ad hoc inferential mechanisms to the knowledge represented in the guideline. For instance, Asbru [15] and GLARE [16] adopt temporal-reasoning algorithms for temporal consistency checking, useful both for GL acquisition, and for simulation purposes. In GLARE, costs and resources required by the various GL actions can be collected, and the "admissible" paths in the GL (e.g. paths not exceeding a prefixed cost) can be identified on the basis of this result. Several systems [17] apply some controls for checking the well-formedness of the acquired GL, e.g. as regards name and range

* Corresponding author. Tel.: +39 0131360174; fax: +39 0131360198.
E-mail address: terenz@mfn.unipmn.it (P. Terenziani).

0933-3657/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.artmed.2009.09.003

checking of the actions of a GL and of its attributes (which must match specific standards), or as regards the adherence to several logical design criteria, such as the fact that alternative arcs may only stem from decisions.

However, two major drawbacks of the conventional approach can be outlined as follows (for a more detailed discussion see Section 2.2):

(i) every class of properties to be checked, for every GL (possibly with the exceptions of purely syntactical properties), requires the definition of an ad hoc verification software module. The analysis of an additional class of properties thus requires an additional effort by the programmers who are in charge of verification;
(ii) the verification process is not conceived as a flexible and incremental one: all properties to be verified must be known a priori, in order to let the verification software modules be developed before the verification process starts. Additional relevant classes of properties suggested by the already obtained results cannot be easily taken into account (due to the issue discussed in point (i)).

More generality and flexibility are therefore needed in guideline property verification. Generality is one of the main achievements of the theorem provers and model checkers developed within the automatic verification community [18]. Therefore, integration between the "physician-oriented" way of coping with clinical guidelines supported by the guideline management systems on one side, and the generality of verification techniques, on the other side, can provide fruitful results. Such an integration has started to be explored only quite recently within the Medical Informatics community. The adoption of theorem proving techniques has been first proposed within the Protocure European project starting in 2003 [2,19]. As an alternative of the theorem proving methodology, the adoption of model checking techniques has been first proposed few years later in the Protocure project [20] and in our GLARE project [21–23], mainly motivated by the simplicity and efficiency of model checking techniques with respect to the theorem proving approach [24]. In this paper we elaborate on the ideas first sketched in [21–23], extending and systematizing such an initial proposal (as discussed in Section 8, where we also explore in-depth the main differences between our approach and Protocure's one).

In particular, our paper focuses, on one side, on the knowledge representation and methodological issues (which are typically the main interest of *AI researchers*), and, on the other side, on usability issues (which are more interesting from the medical point of view) by analyzing which properties of the guidelines can be verified, and when. More specifically, the paper main contributions are the following:

(i) first, as a motivation for our approach, we propose an in-depth analysis of when the verification capabilities we provide can be used within the GL life-cycle, and a general overview of the different types of properties that can be verified (i.e., what can be verified);
(ii) second, we provide a general methodology to integrate verification capabilities within a GL management system. Specifically, we propose a modular approach in which a computerized GL management system is loosely coupled with a model-checker via a translator, which maps any GL expressed in the formalism of the computerized GL management system into the formalism of the model-checker. In such a way, the advantages of adopting a GL management system from one side, and a general-purpose model-checker on the other side are retained and combined. In particular, once the mapping has been defined, any class of properties that can be formalized in the logic of the model-checker can be easily verified, without requiring the definition of a new verification software module from scratch. This obviously facilitates a real interaction between the physician examining the GL and the system itself. Thanks to its modularity, such an approach can be easily implemented, since it does not require any modification to either the computerized GL management system or the model-checker;
(iii) third, we show how such a general approach can be instantiated. Although our proposal is mostly application-independent, as a proof of concept, we are currently integrating within the system GLARE [7] a verification tool which models a GL in Promela, the specification language of the model-checker SPIN [25], and verifies the GL properties to be checked by formalizing them as Linear time Temporal Logic (LTL) formulas. In particular, one of the contributions of the work relies in the analysis of how a GL can be represented in a process-based language such as Promela;
(iv) fourth, we refine the discussion about the different types of properties proposed in item (i), showing how they can be expressed using LTL. We also propose an application to the verification of the guideline about ischemic stroke as a concrete example.

The paper is organized as follows. In Section 2, we introduce our general goals and methodological choices. In Section 3 we summarize the main features characterizing GLARE, which will be needed to present our verification approach, and in Section 4, we briefly introduce the model-checker SPIN. In Section 5, we specifically present our implementation of model checking for verification in GLARE. In Section 6, we show several properties than can be checked during the GL life-cycle, classified as in Section 2. Section 7 contains a more extensive verification example, conducted on a real world GL. Section 8 is devoted to comparisons with related work. Finally, Section 9 contains our concluding remarks and future research directions.

2. General goals and methodology

In this section, we first show the advantages of adopting property verification throughout the computerized GL life-cycle, and then introduce our methodological approach to GL verification.

2.1. Using verification throughout the computerized GL life-cycle

It is important to recognize that, in the computerized GL life-cycle, different phases can be distinguished, and different actors play an important role. Specifically, we single out three main phases (namely (1) design and acquisition, (2) contextualization, and (3) execution), and we highlight how verification can be fruitfully exploited in each phase. As a result of such an analysis, different classes of properties are identified. Such classes will be further on elaborated, discussed and exemplified in Section 6.

2.1.1. Design and acquisition

GLs are usually defined by a national or international committee of specialists, and can be acquired into a computer-based system, usually through a cooperation between some specialists and some knowledge engineers. In such a phase, verification through model checking is useful in order to take into account at least two different classes of properties, namely structural properties and medical validity properties. In particular:

(i) *Structural properties* concern the existence of the appropriate clinical requirements. These properties regard the actions,