Information Systems 000 (2018) 1–13

Contents lists available at ScienceDirect

Information Systems

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

Checking process compliance against natural language specifications using behavioral spaces

Han van der Aa[a,*], Henrik Leopold[a], Hajo A. Reijers[a,b]

[a] Department of Computer Sciences, VU University Amsterdam, Faculty of Sciences, De Boelelaan 1081, Amsterdam, 1081HV, The Netherlands
[b] Department of Mathematics and Computer Science, Eindhoven University of Technology, PO Box 513, 5600MB, Eindhoven, The Netherlands

ARTICLE INFO

Article history:
Received 14 December 2016
Revised 19 July 2017
Accepted 25 January 2018
Available online xxx

Keywords:
Business process analysis
Compliance checking
Natural language processing
Ambiguity

ABSTRACT

Textual process descriptions are widely used in organizations since they can be created and understood by virtually everyone. Because of their widespread use, they also provide a valuable source for process analysis, such as compliance checking. However, the inherent ambiguity of natural language impedes the automated analysis of textual process descriptions. While human readers can use their context knowledge to correctly understand statements with multiple possible interpretations, automated tools currently have to make assumptions about their correct meaning. As a result, compliance-checking techniques are prone to draw incorrect conclusions about the proper execution of a process. To provide a comprehensive solution to these reasoning problems, we use this paper to introduce the concept of a *behavioral space* as a means to deal with behavioral ambiguity in textual process descriptions. A behavioral space captures all possible interpretations of a textual process description in a systematic manner. Thus, it avoids the problem of focusing on a single, possibly incorrect interpretation. We use a quantitative evaluation with a set of 47 textual process descriptions to demonstrate the usefulness of a behavioral space for compliance checking in the context of ambiguous texts.

1. Introduction

Non-compliance represents a risk for many organizations. According to a recent study by Thomson Reuters, non-compliance may even represent a possible cause of bankruptcy, also for the so-called "behemoths" in the financial sector [1]. Recognizing the risk that is associated with non-compliance, organizations in a wide range of domains are stepping up their spending in order to ensure their compliance with laws, regulations, and procedures. In this context, automated compliance checking techniques play a crucial role thanks to their ability to automatically identify compliance violations [2,3]. For this reason, numerous approaches have been developed to perform this task (cf. [4–7]). What these compliance-checking techniques have in common is that they rely on a structured specification of allowed behavior, for example in the form of *process models* or *business rules*. As a result, these techniques ignore the wealth of information that is contained in less structured forms of process documentation, such as textual process descriptions [8].

While the relevance and widespread use of text documents as a source for process analysis has been emphasized in various contexts [9–12], the inherent ambiguity of natural language presents a considerable challenge to compliance-checking techniques. For example, a simple natural language statement such as "*in parallel to the latter steps*" leaves room for interpretation. Due to this statement's ambiguity, it is generally impossible to infer with certainty whether "*latter*" refers to the preceding two, three, or even more activities mentioned in the textual description. In prior work, text-to-process model generation techniques have circumvented this problem by introducing interpretation heuristics [9,13,14]. In this way, these techniques obtain a single process-oriented interpretation of the text, in spite of the presence of ambiguous sentences. This interpretation, however, contains assumptions on the correct interpretation of essentially undecidable ambiguity issues. So, there is always the risk that the derived interpretation conflicts with the proper way to execute the process. As a result, the focus on a single, assumed interpretation can lead to incorrect and, thus, untrustworthy compliance-checking results.

To provide a rigorous solution for the reasoning problems caused by ambiguous natural language statements, we introduce a novel concept which we refer to as a *behavioral space*. A behavioral space precisely captures all possible behavioral interpretations of a textual process description. The behavioral space clearly defines

* Corresponding author.
E-mail addresses: j.h.vander.aa@vu.nl (H. van der Aa), h.leopold@vu.nl (H. Leopold), h.a.reijers@vu.nl (H.A. Reijers).

https://doi.org/10.1016/j.is.2018.01.007
0306-4379/

After a claim is received, a claim officer reviews the request and records the claim information. The claim officer then validates the claim documents before writing a settlement recommendation. A senior officer then checks this recommendation. The senior officer can request further information from the claimant, or reject or accept the claim. In the former case, the previous steps must be repeated once the requested information arrives. If a claim is rejected, the claim is archived and the process finishes. If a claim is accepted, the claim officer calculates the payable amount. Afterwards, the claims officer records the settlement information and archives the claim. In the meantime, the financial department takes care of the payment.

Fig. 1. Exemplary description of a claims handling process.

which behavior is within and which behavior is outside any reasonable bounds of interpretation. By using behavioral spaces for compliance checking, we avoid the need to impose assumptions on the correct interpretations of ambiguous natural language texts. Therefore, compliance checks based on behavioral spaces provide trustworthy results: They avoid the risks associated with the selection of incorrect interpretations.

The remainder of the paper is structured as follows. Section 2 motivates the problem of reasoning under behavioral ambiguity in textual process descriptions. In Section 3, we introduce the notion of a behavioral space to capture behavioral ambiguity. Section 4 describes how a behavioral space can be generated from a textual process description. We show how to perform compliance checks using a behavioral space in Section 5. Then, Section 6 introduces a semi-automated pruning technique that can be used to effectively reduce the uncertainty in compliance-checking results. In Section 7, we demonstrate the usefulness of behavioral spaces and our proposed pruning technique through a quantitative evaluation using real-world data. Section 8 discusses streams of related work. Finally, we conclude the paper and discuss directions for future research in Section 9.

2. Behavioral ambiguity in textual process descriptions

In this section, we illustrate the problem associated with compliance checking of process behavior against textual process descriptions. The key challenge in this context is the inherent ambiguity of natural language. *Ambiguity* in natural language refers to a type of uncertainty in which several interpretations of the same text are plausible. For example, the sentence "*I saw a man on the hill with a telescope*" can have at least five plausible interpretations. These interpretations vary, among others, on who is on the hill (*I* or the *man*) and on the possessor or the location of the telescope (*I*, the *man*, or it is *on the hill*). In certain situations, the correct interpretation of an ambiguous statement can be clear from the context in which it is used, whereas in other situations even context cannot help to resolve ambiguity.

The goal of compliance checking is to determine if some observed behavior (i.e. a sequence of performed activities) conforms to the allowed behavior described by a process specification, i.e. a textual process description. Therefore, in a compliance-checking context we are particularly concerned with ambiguity related to the allowed process behavior described in a text, which we shall refer to as *behavioral ambiguity*. Behavioral ambiguity occurs when statements about the relations that exist between process steps can be interpreted in different ways. We illustrate the problem of behavioral ambiguity through the simplified description of a claims handling process, as presented in Fig. 1. The description uses typical patterns to describe ordering relations, as observed in process descriptions obtained from practice and research [9]. At first glance, the description from Fig. 1 may appear to be clear. However, on closer inspection, it turns out that the description does not provide conclusive answers to several questions regarding the proper execution of the described process. For instance:

Q1. Is it allowed that the claims officer records the claim information *before* reviewing the request?
Q2. Does it suffice for the claim officer to rewrite the settlement recommendation in case additional information has been requested?
Q3. Can the financial department start paying the claimant while the settlement information is still being recorded?

Based on the information provided in the textual description, these questions are not clearly decidable. This lack of decidability results from two forms of behavioral ambiguity: type ambiguity and scope ambiguity. *Type ambiguity* occurs when a textual description does not clearly specify the type of order relationship between two activities. For instance, the relation between the "*review request*" and "*record claim information*" activities in the first sentence is unclear. The term "*and*" simply does not allow us to determine whether these activities must be executed sequentially or whether they can be executed in an arbitrary order (Q1). *Scope ambiguity* occurs when statements in a textual description underspecify to which activity or activities they precisely refer. This type of ambiguity particularly relates to repetitions and parallelism. For instance, the statement "*the previous steps must be repeated*" does not clearly specify which activities must be performed again (Q2). Similarly, the expression "*in the meantime*" does not define when the financial department can start performing its activities (Q3).

As a result of such ambiguities, there are different views on how to properly carry out the described process. When deriving a single structured interpretation from a textual process description, as done by process model generation techniques (cf. [9,13,14]), there is always the risk that a derived interpretation conflicts with the proper way to execute the process. The focus on a single interpretation can, therefore, lead to wrong conclusions when reasoning about a business process. This can, for instance, result in a loss of efficiency by not allowing for parallel execution where possible (Q3). Furthermore, it can result in non-compliance with regulations, for example, by failing to impose necessary ordering restrictions (Q1) or by not repeating all of the required steps when dealing with the receipt of new claim information (Q2).

To avoid the problems associated with using an assumed interpretation, automated reasoning techniques should take into account all reasonable interpretations of a textual process description. Therefore, we use this paper to introduce the concept of a *behavioral space*. A behavioral space allows us to capture the full range of possible semantics that can be conveyed by textual descriptions in a structured manner. As such, it provides the basis to correctly reason about compliance to described processes.

Please cite this article as: H. van der Aa et al., Checking process compliance against natural language specifications using behavioral spaces, Information Systems (2018), https://doi.org/10.1016/j.is.2018.01.007