Measuring the Progress of Reference Model-Based Business Process Modeling

Agnes Koschmider\textsuperscript{1,2}, Jose Luis de la Vara\textsuperscript{3}, Juan Sánchez\textsuperscript{3}

\textsuperscript{1}Department of Computer Science
University of Pretoria
Pretoria 0002, South Africa
akoschmider@cs.up.ac.za

\textsuperscript{2}Institute of Applied Informatics and Formal Description Methods
Karlsruhe Institute of Technology, Germany

\textsuperscript{3}Centro de Investigación en Métodos de Producción de Software
Universidad Politécnica de Valencia, Spain
{jdelavara,jsanchez}@pros.upv.es

Abstract: Business process modeling plays a major role in industry and academia. As a result, current research focuses on its assistance. Business process modeling is a time-consuming task and its effort is hard to predict, and a plethora of concepts exists for facilitating and improving it. A concept that has not been studied yet is prediction of the remaining effort of business process modeling. In this paper we tackle this issue and suggest a solution for measuring the progress of business process modeling. Our solution is based on the assumption that a reference model is given representing a best practice of a business process. In addition, a set of process variants exists for that reference model. Under these conditions, we propose two measurement approaches that base on the longest common subsequence algorithm and determine the effort required to complete business process modeling. The main benefit of the approaches is the possibility of effectively structuring and monitoring the development of business process modeling projects.

1 Introduction

The importance of business process modeling (regarded as the process of creating business process models) nowadays is undeniable. It plays a major role in many fields both in industry and academia [In09], such as business process management and information system development. Business process models are used in organizations for documentation issues, for decision support about process development and design, for monitoring of process executions, and for analysis of technology support [Ag04].

Quality and ease of modeling are two of the needs that have been acknowledged for business process modeling [In09]. As a result, much research and many initiatives focus on assistance during its development. For example, existing works try to facilitate and improve business process modeling by applying process mining [Aa07], by
defining metrics for business process models [Me08] and by providing tools that facilitate modeling and guarantee correctness [HKL08, MM09]. Consequently, new ideas can be studied for improving and facilitating business process modeling, and the impact and benefits of their application must be analyzed.

An idea that has not been studied for business process modeling is the application of progress measurement. Progress measurement of processes is a common activity in many fields, but it has not been taken into account for business process modeling yet. Progress measurement is usually related to project management, and allows people to precisely inform about process status and the amount of work that has been carried out at a given moment. Therefore, prediction of work completion is facilitated.

The aim of this paper is to adopt the idea of progress measurement in order to facilitate business process modeling. The solution that is presented assumes two conditions. Firstly, it assumes the existence of a reference model representing a best practice of a business process (e.g. an ITIL model). This assumption can always be met thanks to existing techniques for inferring reference models [LRW08]. Secondly, the solution assumes the existence of a set of process variants for the reference model.

Under these assumptions, two measurement approaches are presented. The strict measurement approach assumes that the process designer requests a full compliance of a business process model under construction with a reference model and its process variants. In this case, the approach estimates the modeling progress based on the number of operations that are necessary in the business process model under construction to comply with the reference model. The relaxed measurement approach assumes that the process designer is looking for a close (non-full) compliance with the reference model. It is a relaxed version of the prior approach.

Both approaches base on the longest common subsequence (lcs) algorithm, for which efficient versions exist even for more than two input sequences [Ba07]. By applying the lcs algorithm for our purpose, we can detect the longest, not necessarily contiguous, sequence of activities that two business process models have in common. Inputs for the algorithm are activity labels of business process models.

Application of progress measurement will be shown for a modeling support tool. The recommendation-based editor [HKL08] suggests users appropriate business process models or fragments that are similar to the ones that they intend to create. The application of the measurement approaches in this context might help users to choose business process models that will allow them to finish modeling faster. Another advantage is that users can request the compliance with a reference model, and compliance can be regarded as related to the quality characteristic of completeness of a business process model (provision of complete information) [BGR06].

The paper is organized as follows. Section 2 describes the measurement approaches. Section 3 shows application of progress measurement in the recommendation-based editor. Section 4 discusses about the solution that is presented. Section 5 presents related work. Finally, Section 6 summarizes our conclusions and future work.
2 Measurement Approaches

The measurement approaches base on two assumptions in order to predict the effort that is required to complete business process modeling.

**Assumption 1.** A reference model is given that describes a best practice.

**Assumption 2.** A set of process variants exists for this reference model and are stored in a repository.

The reference model may be, for example, a SCOR, ITIL or CobiT model. In case that no reference model is given, [LRW08] have suggested a heuristic method to determine a reference model from existing business process models. Consequently, assumption 1 can be derived, while 2 must be given.

The measurement approaches do not consider the behavior of activities business process models (branching, merging, fork, join, etc.). Instead they take activity labels (names) as input. In the rest of Section 2 we treat activity labels as letters. For example, “fill customer information” or “choose extras” are represented as “A” or “B” for the sake of understandability and simplicity.

[KSR10] found out that the semantics of process element names is more important for users than syntax in tools that assist users in their modeling task. The reason for that finding is that structuring of process elements is easier than creation of business process models (a model recommendation can be regarded as an inspiration that can individually be adapted and structured).

Certainly, process elements can be labeled with different styles (e.g. verb-object or action-noun) or using different vocabulary (thus synonyms, homonyms or hypernyms can arise), which hampers the reduction to letters. An initial (semi-) automatic approach has been proposed to detect the labeling style and to uniform labels to a preferred style [LSM09]. Unambiguities in process element labels can also be detected using the approaches presented in [EKO07, HKL08]. Therefore, the substitution of process labels with letters is justified, assuming that techniques to detect unambiguity are applied and inhomogeneous labeling styles are unified prior to the usage of the measurement approaches.

Figure 1 shows the scenario for application of our measurement approaches. A process designer is creating a business process model that should comply with a reference model. She is also using a repository of business process models from which she obtains recommendations for creating her model. The modeling progress is calculated on the basis of the compliance of the process being designed (i.e. the business process model under construction) with its reference model and a set of process variants. The process variants, which are stored in the repository, are variants of the reference model. The process designer can decide between the strict and the relaxed measurement approaches to calculate the modeling progress.
The intention of both approaches is to find the smallest set of operations (insertions and/or deletions) that is required to complete reference model-based business process modeling on the basis of lcs. For example, let “ABCDMN” be a reference model (abstractly) and let “ADY” be a process being designed, then the result of lcs for them is “AD”. If lcs is not empty, then subsequently the diff method determines the number of change operations that are necessary to transform the process being designed into a business process model that is compliant with the reference model. Based on the change effort, the remaining work that is required to obtain a compliant business process model can be calculated.

The following abbreviations are used in the rest of Section 2. ‘R’ refers to the reference model with which a business process should comply. The process being designed is referred to as ‘P’. The set of process variants that are stored in a repository are referred to as ‘{PV}’. Each process variant is referred to as ‘PVX’, where ‘X’ is a number that identifies it. These abbreviations have been included in Figure 1, which is also used as an example to show the measurement approaches.

The next subsections present the strict and the relaxed measurement approaches in detail. The strict measurement approach assumes a full compliance with the reference model and its process variants, whereas the relaxed one relieves this requirement. The process designer can decide what kind of compliance she desires.

2.1 Strict Measurement Approach

Development of the strict measurement approach consists of eight steps.

1) Find dependency between P and R and between P and {PV}
The intention of the first step is to ensure that the activities of P are related to R or to any process variant of {PV}.
Input: P, R, \{PV\}

Based on this input, lcs between R and P (lcs_1) is determined. Subsequently, lcs between P and \{PV\} (lcs_2) is computed. The calculation order of lcs_1 and lcs_2 does not matter. lcs_2 could be determined first and then lcs_1. In case that neither lcs_1 nor lcs_2 have a solution (i.e. lcs_1 is “ ” and lcs_2 is “ ”), then modeling progress cannot be calculated. No solution implies no dependency between the business process models exist. Progress measurement continues if at least a solution is found. In addition, the results of lcs_1 and lcs_2 do not need to be identical. Shorter results mean that more change operations are required to transform a business process model into another and to finally provide full compliance. The number of change operations affects modeling progress.

Output: lcs_1(P,R), lcs_2(P,\{PV\}) (i.e. lcs of P and each process variant of \{PV\})

Result for Figure 1: lcs_1(P,R) is “AD”, lcs_2(P,PV1) is “A”, lcs_2(P,PV2) is “Y”, lcs_2(P,PV3) is “AD”, and lcs_2(P,PV4) is “AD”.

2) Find dependency between R and \{PV\}

The intention of the second step is to find a dependency between R and \{PV\}. This step is also a preparation operation for step 3.

Input: R, \{PV\}

Based on this input, lcs between R and \{PV\} (lcs_3) is determined. The following steps of the strict measurement approach are only applicable for process variants where lcs_3 has a solution. The repository can include process variants that have activities in common with P but not with R (e.g., PV2 in Figure 1, which includes “Y”). Such process variants are not further considered because they do not comply with R.

Output: lcs_3(R,\{PV\})

Result for Figure 1: lcs_3(R,PV1) is “ABC”, lcs_3(R,PV2) is “ ” (empty), lcs_3(R,PV3) is “ABCD”, and lcs_3 (R,PV4) is “ABCD”. Consequently, PV2 is not further considered.

3) Find the “highest” compliance of \{PV\} with R

The process variants have a different compliance degree with the reference model. A full compliant process variant includes all activities of the reference model. The intention of this step is to find the process variant(s) that incorporates the highest compliance with a reference model. Parsing of the business process model or process variants is performed on the basis of the Refined Process Structure Tree [VVK09].

Input: lcs_3(R,\{PV\}) where lcs_3 is not “ ”

Based on this input, the absolute lcs (lcs*, longest lcs_3) is calculated.

Output: lcs*(R,\{PV\})

Result for Figure 1: lcs* is “ABCD”.

4) Determine the distance between P and the “highest” compliance of \{PV\} with R

For the strict measurement approach, we assume that the process designer desires full compliance of P with R and \{PV\}. The intention of this step is to determine the operations (specified as ‘[insertion or deletion, position, activity label]’, where insertions are represented with the symbol “+” and deletions with the symbol “-”) that are nec-
ecessary to extend P with lcs*. Then P incorporates the highest compliance degree of \{PV\} with R.

\textit{Input:} P, lcs*

Based on this input, P is extended with the activities that are part of lcs* but not of P. It is then denoted as P'. In case that no operations were necessary, then P and P' would be identical.

\textit{Output:} P'

\textbf{Result for Figure 1:} insertions of “B” at the second position of P (\([+,1,“B”]\)) and of C at the third position (\([+,2,“C”]\)) are necessary. P’ is “ABCDY”.

5) \textbf{Apply lcs* for \{PV\}}

The intention of this step is to determine the activities that are part of \{PV\} but not of lcs* or P’. As P’ incorporates lcs*, each process variant of \{PV\} is extended so that lcs* holds for all of them. This step is a preparation operation for the step 6.

\textit{Input:} P’, lcs*, \{PV\}

Based on this input, each process variant is extended so that lcs* holds. The result is denoted as \{PV’\}.

\textit{Output:} \{PV’\}

\textbf{Result for Figure 1:} PV1’ is “ABCDF”, PV3’ is “ABCD” and PV4’ is “ABCDEFG”

6) \textbf{Determine the difference (diff) between P’ and \{PV’\}}

The intention of this step is to determine the smallest set of operations for turning P’ into \{PV’\}, and consequently to determine all the activities that are part of in \{PV’\} but are not yet of P’. This step guarantees full compliance of P’ with \{PV’\}.

\textit{Input:} P’, \{PV’\}

Based on this input, P’ is extended on the basis of \{PV’\}. The result is denoted as P’’. This operation excludes activities that are unique in P’ (in the example, “Y”). We assume that the process designer is not modeling an identical process variant thus she is free to insert as much unique activities as she likes in the process being designed.

\textit{Output:} P’’

\textbf{Result for Figure 1:} two insertions (diff(P’,PV1’)) is 1, \([+,5,”F”]\); diff(P,PV3’) is 0; diff(P,PV4’) is 1, \([+,5,”E”]\)) are required to make P’ fully compliant with \{PV’\}. P’’ is “ABCDYEF”.

7) \textbf{Determine the difference (diff) between P’’ and R}

This step determines the operations that are necessary in P’’ to fully comply with R.

\textit{Input:} P’’, R

Based on this input, P’’ is extended with the activities that are part of R but not yet of it. The result is denoted as P’’’.

\textit{Output:} P’’’

\textbf{Result for Figure 1:} P’’’ is “ABCDYEFMN”, which results from two insertions in P’’ (\([+,7,”M”]\), \([+,8,”N”]\))

8) \textbf{Calculate the progress}

The result of this step is an estimation of the percentage of effort that has already been carried out in P. It is calculated on the basis of the following formula:
Modeling_Progress = \frac{\text{length} \left( \text{lcs}_1(P,R) \right)}{\text{length} \left( \text{lcs}_1(P,R) \right) + \text{number of operations}}

For the example, Modeling_Progress = \frac{2}{2+6} = \frac{2}{8} = 0.25 , where length(\text{lcs}_1 (P,R)) is the number of activity labels of \text{lcs}_1 (“AD”) and the number of operations corresponds to the operations determined in steps 4, 6 and 7 (2 operations in each step). This modeling progress result means that the process designer has modeled 25% of her business process model (with respect to full compliance with R) in P, assuming that she will use \{PV\}. In the next subsection we discuss the relaxed measurement approach, which eases the compliance degree.

### 2.2 Relaxed Measurement Approach

The relaxed measurement approach assumes that a process designer does not aim to obtain full compliance of P with R and \{PV\}, but she is satisfied with a close (non-full) compliance. The relaxed version results from the following modifications:

- The process designer can consider process variants that are not compliant with R but have common activities with P. This eliminates step 2 of the strict measurement approach. Referring to Figure 1, PV2 would be further considered.

- The process designer can specify her compliance preference to \{PV\}, for example, based upon the frequency of activities in \{PV\}. The frequency is shown to the process designer in descending order where the most frequent activities (e.g., F occurs twice, E and Z once) are shown on the top. For instance, in Figure 1 the process designer may desire activity E in her business process model instead of F even if F occurs more times.

- No full transformation of P” into R is required. The process designer decides what activities of R she desires to include in her business process model.

The advantage of this approach is that process designers can browse the set of remaining activities and opt for specific ones. For example, on the basis of Figure 1 and referring to the steps of the strict measurement approach, if a process designer (1) desired no compliance with \{PV\} (P’ and P” would be identical; no operations would be determined in step 6), and (2) decided not to include “M” in her business process model (no full compliance with R is required; just an operation would be determined in step 7), then Modeling_Progress = \frac{2}{2+3} = \frac{2}{5} = 0.4 , where three operations of the strict measurement approach would not be necessary. Therefore, the process designer would have modeled 40% of her business process model (for non-full compliance with the R) in P.

In summary, the relaxed measurement approach can be regarded as the upper bound of modeling progress and the strict one as the lower bound. In this sense, if a process designer decided not to comply with \{PV\} and not to include more activities of R (in step 7), then the modeling progress would be 100% (maximum upper bound).

224
3 Progress Measurement in a Recommendation-Based Editor

A recommendation-based editor for business process modeling [HKL08, KHO10] is used as an example of application of the measurement approaches. The editor is characterized by the use of a recommendation engine that suggests business process models or model fragments from a process repository in order to help process designers complete a business process model under construction.

The process repository is populated with business process models that are stored in the repository after their completion. These business process models are then accessible and thus can be re-used. To retrieve the models, the recommendation-based editor provides: 1) a query interface that allows users to request business process models or fragments that are of interest to them (e.g. by providing the name of a model and its first and last elements), and; 2) a recommender component that automatically suggests appropriate process model fragments for model completion. The decision for a recommendation mainly depends on the score value (Figure 2).

Figure 2: Progress measurement included in a modeling support tool

The score value specifies the linguistic similarity between business process models and the query input. If a user is interested in a recommendation, then the graphical view of the business process model can be opened (button “Open graphical view”).

If progress measurement is not taken into account, users have to decide by themselves which recommendation would result in the fastest completion of a business process model. In Figure 2 the first two recommendations have a similar score value. The progress value is an additional decision support in favor of a recommendation.

Another advantage of the integration of progress measurement into this modeling tool is to inform users about the compliance of the recommendations with a reference model. The score value does not consider the order of activities in contrast to the measurement approaches. A high score may result from processes having a low compliance degree. When clicking on the button “Modeling progress details”, all the operations that are necessary to complete a business process model are listed.

More concrete details about how modeling progress is measured in the recommendation-based editor are out of the scope of this paper.
4 Discussion

Discussion about the measurement approaches focuses on their effectiveness and on further application areas of progress measurement of business process modeling.

4.1 Effectiveness of the Measurement Approaches

As explained in Section 2, our measurement approaches assume two conditions, where the second one must be fulfilled: a set of process variants of a reference model (with which a process designer wants a business process model to comply) are given. Companies that use the ITIL standard usually customize its models and create process variants. Thus, our assumptions are realistic.

The measurement approaches provide an estimation of the required effort to complete business process modeling. If a process designer had a completely different business process in mind (that only closely resembled the reference model and its process variants), then the progress value could only be regarded as an approximation.

One can criticize the effectiveness of our measurement approaches due to the limited focus on activity labels. Even if the approaches do not consider the behavior of business process models, they consider the order of activities. For this, the diff method reorders the remaining (missing) activities needed so that a business process model under construction is analog to the reference model.

4.2 Further Application Areas of Progress Measurement

Progress measurement could be applied in other business process modeling contexts. For instance, in collaborative, on-line business process modeling the measurement approaches might be used to monitor the development of business process modeling projects. In such collaborative settings, process designers store their models in open repositories and make them available to others. Given the large amount of business processes, it should be easy to determine a reference model (using the technique of [LRW08]) to measure the modeling progress and monitor collaborative activities.

The measurement approaches might also be incorporated in approaches that calculate the similarity or compliance between business process models based on their behavior (some examples are shown in Section 5). Existing works that identify the compliance degree disregard the linguistics of activities and just focus on structure and behavior of business processes. Our measurement approaches are complementary to behavior-based compliance measurement.

Finally, progress measurement could be applied in organizational modeling approaches that model business processes from existing information and/or other models (e.g. [Sc00, VSP08]). On the basis of the proportion of elements that are present in a business process model, modeling progress might be estimated.
5 Related Work

We are not aware of any work that has addressed progress measurement of business process modeling. Nonetheless, five streams of work are related to this paper: (1) progress measurement of processes, (2) application of lcs, (3) measurement of characteristics of business process models, (4) compliance of business processes, and (5) similarity between business process models.

Progress measurement of processes is a common practice in many fields such as project management [CI06], software engineering [LB06], industrial engineering [BW04] and construction [Zh09]. It is used to improve knowledge about process development and status. These works confirm that progress measurement is an important and common activity. However, the nature of the processes of these fields is different from business process modeling (e.g. in its description). Therefore, their measurement approaches cannot be directly applied to business process modeling.

Use, application and adaptation of lcs is common in many research fields such as image pattern recognition [LSY89] and bioinformatics [JP04]. Although lcs has been used in business process management [GCC09], no work has applied it to measure the progress of business process modeling.

Within business process management, several works deal with measurement of different characteristics of business process models in order to evaluate their usefulness (e.g. [Me08]). These characteristics are complexity, understandability, quality, entropy, density, cohesion and coupling [Go10]. The main difference of our measurement approaches is that they do not only aim to improve knowledge about and quality of business process models, but also to accelerate process designers’ work.

Some works address compliance of business process models with requirements such as regulations [Kh08], data objects lifecycles [KRG07] and goals [KK97]. Compliance can be measured based upon business process management platforms [LSG08] or behavioral profiles [We10]. The reference models that are used in our measurement approaches can be regarded as another type of requirements. Our measurement approaches determine compliance on the basis of activity labels, but they can easily be applied to behavioral-based approaches.

Research on similarity between business process models (e.g. [MAW08, DDG09]) is also related to this paper. In [GCC09], similarity to reference models is addressed. Most of the works focus on measuring similarity between business process structures, and none address measurement of modeling progress. Nonetheless, their methods for similarity measurement could be adapted for the latter purpose.

Last but not least, prediction of completion of business process execution [Aa09] is a visionary proposal that is based on features of navigation systems. Since completion of a journey can be predicted, completion of business process execution could also be done. However, the way to carry out the prediction has not been determined yet.
6 Conclusions and Future Work

Business process modeling is essential in many fields nowadays, and much research and many initiatives have been proposed in order to facilitate and improve its development. Nonetheless, new ideas can be analyzed and their benefits must be studied.

This paper has proposed progress measurement as a new practice for supporting business process modeling, whose application has not been previously analyzed. As a result, two measurement approaches have been presented for reference model-based business process modeling. Both approaches calculate the modeling progress based on the compliance of activity labels and use the lcs algorithm for this purpose.

Progress measurement can improve business process modeling from two perspectives. From a perspective of ease of modeling, business process modeling is improved by providing information about the amount of work that has been carried out and/or that is still needed. Process designers' work might also be accelerated. From a perspective of quality, business process modeling is improved by providing information about the completeness of a business process model. Application of progress measurement in a modeling tool and discussion about the paper have also been presented.

As future work, we plan to apply progress measurement in actual settings and projects, to analyze the efficiency of the recommendation-based editor when calculating modeling progress, and to further study application of progress measurement in other methods for and applications areas of business process modeling.

Acknowledgements. This work has been developed with the support of the Spanish Government under the project SESAMO TIN2007-62894 and the program FPU AP2006-02324, and has been co-financed by FEDER.

References


