Decomposition of business process models into reusable sub-diagrams

Piotr Wiśniewski

AGH University of Science and Technology, Department of Applied Computer Science, al. Mickiewicza 30, 30-059 Krakow, Poland

Abstract. In this paper, an approach to automatic decomposition of business process models is proposed. According to our method, an existing BPMN diagram is disassembled into reusable parts containing the desired number of elements. Such elements and structure can work as design patterns and be validated by a user in terms of correctness. In the next step, these component models are categorised considering their parameters such as resources used, as well as input and output data. The classified components may be considered a repository of reusable parts, that can be further applied in the design of new models. The proposed technique may play a significant role in facilitating the business process redesign procedure, which is of a great importance regarding engineering and industrial applications.

1 Introduction

Business Process Management stands for a set of techniques which aim to design, analyse, implement and improve organisational processes. Process models can help organisations in visualising processes of the company and give opportunity for optimisation of the process structure. Owing to this, the company may achieve business goals in a more efficient way.

BPMN (Business Process Model and Notation) [1], maintained by the OMG group, is the most widely adopted notation for modelling business processes. BPMN 2.0 is quite complex, as there are many additional documents explaining the notation, such as handbooks and papers devoted exclusively to the pragmatics of the notation usage. In the paper, a brief overview of the guidelines for process modelling as well as various strategies of process design were provided.

According to the research based on industrial applications [2], most of the manually created process models are badly affected by quality issues. Therefore, an automated support in the phase of process design is highly appreciated. The proposed method may be seen as a business process modelling technique based on sub-diagrams generated as a result of a decomposition of existing models. The workflow specification begins with process elements and sub-processes of which the reusable sub-diagrams are created. These component parts are then combined to form a comprehensive process. An overview of the decomposition method is presented in Figure 1.

This paper is organised as follows. In Section 2, an overview of the existing solutions in the area of process model design, as well as their applications, is presented. Section 3 describes in detail the idea of business process modelling and their mathematical representation. The proposed decomposition algorithm is presented in Section 4. Section 5 shows the idea of a component process repository, while in Section 6 the main guidelines of sub-diagrams merge were described.
process models [3]. This is conducted by identifying SESE (Single Entry, Single Exit) structures in a process tree. Milani et al. [4] discuss criteria and heuristics for model decomposition. They also provide an evaluation regarding the influence of these criteria on the understandability of selected models. The decomposition procedure may be supported by detecting names of non-uniformly specified process element in order to enable a faster creation of consistent complex models [5]. Another solution to the discussed problem was presented in [6] where business processes were presented as a hierarchical ontology of business tasks apt for re-composition. Such a description may simplify the dynamic creation of models of collaborative business processes.

This paper is also related to structured and formal process representations which facilitate process modelling, as well as their reconstruction. One of the user-friendly approaches is the spreadsheet-based representation [7] where the set of activities is listed in a table with specified branches and routing conditions. Process models apt for re-composition can be also described in a declarative specification [8], which focuses on constraints and dependencies between activities instead of determining an explicit sequence flow. They can be also represented as a randomly-ordered list of tasks, along with their input and output conditions, which are then used for a constraint-based process composition [9]. Another approach is a semantic model for BPMN specified in the CSP process algebra [10]. This is especially useful for specifying time constraints between process activities.

Process model repositories are an important tool in the area of business process redesign. They can be developed in a form of a database storing relevant process fragments with a reference to the process metadata [11]. This approach leads to an automated generation of complex models. In case of such a composition, the correctness of the generated process model must be ensured. Therefore, it is necessary to mention the identification and evaluation of issues regarding the flow consistency [12] as well as the layout of the generated business process model [13]. The automated process modelling should also include verification whether the created model contains any typical anomalies such as deadlocks, livelocks or the lack of synchronisation [14], [15] or event anomalies [16]. The analysis of the formal aspects of a BPMN model can be also conducted using Alvis Modelling Language [17].

3 Business Process Models

Highly developed software systems are present in many areas of industrial and business applications, which range from human-machine interface systems, through production management, to web services covering a large domain of implementations. To ensure the reliability and efficiency of such systems and simplify their design and development phases, advanced process engineering approaches are in use.

The engineering process usually starts with business process modelling which constitutes a set of techniques used to represent a certain workflow by providing visual diagrams depicting a business process with its dependencies. Thus, a business process model presents the way operations are performed to accomplish the planned objective of an organisation.

An example field of application for business process models is the specification of Manufacturing Execution Systems (MES) [18]. Building a complete modelling framework for the integration of a system technical and functional model, as well as production process model ensures the understandability of the specification for all the participants involved in the implementation process.

3.1 Business Process Model and Notation

Business process models in BPMN notation are represented by diagrams composed of a limited set of graphical symbols [19]. Visualisation of activity flows allows business users and developers to understand processes more clearly. It is a transparent visual tool dedicated to modelling complex processes.

BPMN model elements can be divided into four major groups: flow objects, connecting objects, swim lanes and artefacts. Flow objects represent the principal elements of the process diagram. In this set, three major object types can be distinguished [1]:
- activities (rounded rectangles) which represent a work that must be performed in the process,
- events (circles) which represent incidents occurring at the time of process execution,
- gateways (rhombi) which control the flow of tokens in the process.

There are two kinds of activities: tasks and subprocesses. A task is an atomic action that can be performed by a person or executed as an automated activity. On the other hand, a sub-process is a compound activity representing a complex work, divisible into smaller parts which are defined as a separate process at the lower level.

The set of events covers three types of elements: start events which indicate the beginning of a process with optional triggering conditions, end events that determine the end of a path in a process or subprocess, as well as intermediate events which may happen at any time within a process.

Gateways control the process flow by splitting it into branches and joining them afterwards. The most typical BPMN gateways are data-based, which means that they evaluate an expression of process data, while they cannot decide regarding the flow of the process [20]. Gateways can be exclusive (XOR) which enable only one outgoing sequence flow based on given process data, inclusive (OR) that precede potentially parallel branches, as well as parallel (AND) used to represent concurrent flows executed without any conditions.

A set of the most typically used BPMN flow objects is presented in Figure 2.

---

ITM Web of Conferences 15, 01002 (2017) CMES’17

DOI: 10.1051/itmconf/20171501002
Fig. 2. Set of common BPMN flow objects. Based on [20].

All the flow objects need to be connected using so-called connecting objects of two major types:
- sequence flow which shows the execution order of a process,
- message flow which represents the exchange of messages between process participants.

To distinguish business functions or parts of the system related to certain flow objects, the concept of swimlanes is usually used in BPMN. Business process participants, which can either be entities within an organisation or different collaborators in a process [21] are represented by pools. Sub-partitions of a pool, which represent specific objects or roles, are called lanes. They organise and categorise flow objects, connecting objects and artefacts in a pool. Lanes are present in BPMN diagrams in a form of a rectangle extended either vertically or horizontally along the length of a pool.

Artefacts show additional pieces of information which enable process designers to include more details in a model. Three main types of artefacts were defined in the standard [1]: data objects, groups and annotations.

3.2 Running Example

To illustrate the proposed method, a sample supply process was selected. Its BPMN model is presented in Figure 3. The process goal is to purchase goods in case if the available inventory is below the desired value.

After the start, a warehouse employee checks the inventory: if the available quantity is above the minimal value, the process is ended; otherwise a requester creates a purchase order which is then reviewed by a supervisor. If the created order does not fulfil the requirements, it should be reprocessed by the requester, otherwise the funds are reserved by the accounting department and the supervisor sends the order to the supplier. If the product is not available at the supplier’s stock, a negative response is sent to the supervisor and the purchase order is cancelled. Otherwise, the customer receives an invoice and a packing slip, which need to be recorded before releasing the funds and issuing the payment. As soon as the payment is sent to the supplier, the process is completed.

3.3 Mathematical Model of a BPMN diagram

To use the proposed process decomposition method, it is necessary to transform a BPMN model into a form of a graph. Therefore, a formal model of the process should be determined before the conversion. A simple BPMN model can be mathematically represented as a tuple $\mathcal{P} = (\mathcal{O}, \mathcal{F})$ [23] where:
- $\mathcal{O}$ is the set of flow objects, such as: activities (including task and sub-processes), events and gateways,
- $\mathcal{F} \subseteq \mathcal{O} \times \mathcal{O}$ is the set of sequence flows.

For the decomposition method, a task with boundary events is represented in the formal model as a single flow object having multiple outputs. In addition, the proposed approach includes neither message flows nor artefacts, as the resulting sub-diagrams are generated only for one determined process participant and additional information may differ depending on a specific process model.

3.4 Conversion to a Business Process Graph

Directed graphs are one of the common representation for business process models [24]. Since the proposed decomposition approach considers only the layout of the process model and not its execution, it was possible to generalise the existing model of atomic and complex process graphs. In the proposed structure called business process graph, the existence of loops as well as multiple start and end events is allowed. Therefore, there are no restrictions regarding the cyclicity of the graph and degrees of its vertices. In this case, the resulting graph model may serve as a precise copy of a BPMN diagram.

Based on the formal model presented in Section 3.3, a business process graph can be defined as a connected, directed graph $G_P = (V_O, E_F)$ where:
- $V_O$ is a non-empty set of vertices representing all flow objects in a process,
- $E_F$ is a non-empty set of edges representing all sequence flows in a process.

Transforming the BPMN model into a business process graph consists in listing all process flow objects along with the following properties:
- Object ID and Name (optional),
- Pool/Lane,
- Object Type (task, XOR split gateway, etc.),
- Condition (optional, used for OR and XOR gateways),
- Input/Output Data Entities (optional).

All the objects represented by vertices should be then connected with directed edges which correspond to the sequence flow of the process. A directed edge $e \in E_F$ is a pair:

$$(v_1, v_2) \in V_O$$

where $v_1$ represents its starting point and $v_2$ its endpoint. In the business process graph edges contain also optional information about a condition value, provided in case of alternative flows starting with XOR and OR gateways.
In the next step, it is necessary to define adjacency matrix $D$ of the generated business process graph. Assuming that the analysed process consists of $n$ flow objects, $D$ is a square matrix of size $n \times n$, defined by Formula (2):

$$
D = (d_{ij})
$$

where rows and column indexes $i,j = 1..n$ correspond to the IDs of process objects. Each element of matrix $D$ is assigned to a value according to Formula (3):

$$
\forall v_i, v_j \in V_O: d_{ij} = \begin{cases} 1, & \text{if } (v_i, v_j) \in E_F \\ 0, & \text{otherwise} \end{cases}
$$

Such a matrix stores all the information regarding flow connections between objects in the analysed process.

### 4 Model Decomposition

The process model decomposition algorithm consists in finding the set $S_k$ of all the subgraphs induced by $k$ vertices, where $2 \leq k < n$. In the first step, it is necessary to define appropriate subsets of $k$ vertices. Then, an adjacency matrix of each subgraph is created by deleting all rows and columns from matrix $D$ which do not correspond to vertices for the selected subset. The following subsections present methods for generating sub-diagrams having two and three elements, as well as the concept of a general solution.

#### 4.1 Bi-gram Decomposition

In the most basic example, the process graph is decomposed into all the possible pairs of vertices connected by an edge:

$$
S_2 = \{\{(v_i, v_j), (v_p, v_s)\}: d_{ij} = 1\}
$$

#### 4.2 Tri-gram Decomposition

A more complex case occurs when there is a need to generate sub-diagrams consisting of three process flow objects which can be connected to multiple elements. In such a situation, a component trigram for an object $v$ can be in one of the following forms:

1. $v$ and its two direct predecessors,
2. $v$ with one predecessor and one successor,
3. $v$ and its two direct successors.

Providing the process graph $G_P = (V_O, E_F)$, having $n$ vertices, as well as its incidence matrix $D$, the decomposition algorithm consists in determining a set $S_3$ by performing the following operation for each vertex $v$ of graph $G_P$:

1. Get the sets of all direct predecessors and successors of $v$, denoted as $Pred(v)$ and $Succ(v)$, respectively.
2. Calculate cardinality of the determined sets, setting $n_p$ as the number of predecessors and $n_s$ as the number of successors.
3. If $n_p > 0$ and $n_s > 0$ then add to set $S_3$ all subgraphs induced by a set of vertices:

$$
\{v_p, v, v_s\}
$$

where $v_p \in Pred(v)$ and $v_s \in Succ(v)$.
4. If $n_p \geq 2$ then add to $S_3$ all subgraphs induced by a set of vertices:

$$
\{v_{p1}, v_{p2}, v\}
$$

where $v_{p1}, v_{p2} \in Pred(v)$ and $v_{p1} \neq v_{p2}$.
5. If $n_s \geq 2$ then add to $S_3$ all subgraphs induced by a set of vertices:

$$
\{v, v_{s1}, v_{s2}\}
$$

where $v_{s1}, v_{s2} \in Succ(v)$ and $v_{s1} \neq v_{s2}$.
In the analysed example presented in Figure 3, 21 flow objects were identified. After running the decomposition algorithm for tri-grams 29 different subgraphs were generated. An example subgraph is shown in Figure 4.

![Subgraph Example](image)

**Fig. 4.** Three-element subgraph generated as a result of process graph decomposition.

Properties of the flow objects corresponding to vertices of the generated subgraph are listed in Table 1 while Table 2 presents edges included in the subgraph.

### Table 1. Properties of flow objects.

<table>
<thead>
<tr>
<th>Object ID</th>
<th>Pool/Lane</th>
<th>Object Type</th>
<th>Object Name</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Warehouse</td>
<td>XOR split</td>
<td>Inventory</td>
<td>min</td>
</tr>
<tr>
<td>3</td>
<td>Operator</td>
<td>end event</td>
<td>Available</td>
<td>--</td>
</tr>
<tr>
<td>4</td>
<td>Operator</td>
<td>task</td>
<td>Create PO</td>
<td>--</td>
</tr>
</tbody>
</table>

### Table 2. Properties of sequence flows.

<table>
<thead>
<tr>
<th>Source Object</th>
<th>Dest. Object</th>
<th>Condition Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>yes</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>no</td>
</tr>
</tbody>
</table>

### 4.3 K-gram Decomposition

The general problem can be solved by determining for a given number $k < n$, all $k$-element subsets of $V_P$ and generating connected subgraphs of $G_P$ induced by these subsets. A related solution was presented in [26], however in this paper, a use of constraint programming technique is proposed.

Following input data should be prepared to solve a constraint satisfaction problem:
- decision variable: $k$-element vector $s_k$ representing a subset of process graph vertices,
- constraints: all elements of $s_k$ are different and connected,
- domain: adjacency matrix of graph $G_P$.

### 4.4 Resulting Sub-Diagrams

The generated subgraphs can be retransformed into BPMN models by executing the inversed method presented in Section 3.4 – vertices of each subgraph represent flow objects and non-zero values of matrix $D$ correspond to the sequence flow of the process.

Each of the sub-diagrams can have a different number of inputs and outputs, according to the flow objects from which it is composed. In general, flow objects can be divided into five main groups:
1. Sources: elements that can contain only output sequence flow, such as start events.
2. Sinks: elements that can contain only outputs, such as end events.
4. SEME (Single Entry Multiple Exit): split gateways, tasks with boundary events, subprocesses.
5. MESE (Multiple Entry Single Exit): merge gateways.

Regarding a best practice of BPMN process modelling [2], implicit splits and joins should be avoided and replaced by gateways. Therefore, if the decomposed model is constructed correctly, all its tasks are SESE elements, except those containing boundary events, as described in Section 3.3. Sub-Diagrams that contain at least one SEME and one MESE object represent the MEME type (Multiple Entry Multiple Exit).

Based on the defined groups, it is possible to determine the number of potential inputs and outputs in each of the generated sub-diagrams, by counting the number of ingoing and outgoing edges of each process flow object and comparing it to its group. This number is used to define functions $\sigma_I/O$, described by Formula 8:

$$
\sigma_I = \begin{cases} 
0, & \text{if there are no available entries} \\
1, & \text{if there is one available entry} \\
2, & \text{otherwise}
\end{cases}
$$

Respectively, function $\sigma_O$ is defined for exits. Table 3 presents to which group a sub-diagram belongs depending on values of functions $\sigma_I$ and $\sigma_O$.

Figures 5-7 present three sub-diagrams of different groups generated based on the example BPMN model presented in Figure 3:
1. SESE (Figure 5) – one available entry for “Reserve Funds” task and one for “Send Order” task.
2. SEME (Figure 6) – one available entry for split AND gateway and two available exits, one per each task.
3. MEME (Figure 7) – one available entry for “Reprocess Order” task, multiple available entries for join XOR gateway and multiple available exits for “Review Order” subprocess.

### Table 3. Sub-Diagram groups depending on functions $\sigma_I/O$.

<table>
<thead>
<tr>
<th>$\sigma_I$</th>
<th>$\sigma_O$</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>subprocess</td>
<td>source</td>
<td>source</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>sink</td>
<td>SESE</td>
<td>SEME</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>sink</td>
<td>MESE</td>
<td>MEME</td>
<td></td>
</tr>
</tbody>
</table>
Towards a Sub-Diagram Repository

The decomposition algorithm presented in Section 4 generates all the possible sub-diagrams of \( k \) elements. This solution ensures that every possible configuration of process elements is ready to be reused in new process models. However, some of the generated solutions may potentially cause process inconsistencies, as they do not contain enough information about the process flow.

Let us analyse one of the tri-grams generated based on the example software testing process model which is presented in Figure 8. In this case, the resulting process fragment will not be fully reusable because of the AND gateway without any output element assigned. Taking into consideration the fact, that the process flow description contains information about the gateway type, as stated in Section 3.4, such a tri-gram configuration prevents from executing only one activity after the “Sign Agreement” task which limits the component utility. At the same time, the number of branches after the AND gateway is not specified which makes the sub-diagram underdetermined.

To ensure the correct repository creation without omitting any relevant process sub-diagrams, a user-based component validation was proposed. The operator who performs the decomposition is presented a set of potentially inconsistent component models. The selected sub-diagrams need to satisfy at least one of the following conditions:

1. Presence of a split gateway without any output branch.
2. Presence of a join gateway without any input branch.
3. More than two swimlanes considered.

In the next step, the component processes are classified considering two criteria: the potential number of inputs and outputs, as well as the diagram similarity. The former is used for grouping models according to the values of functions \( \sigma_I \) and \( \sigma_O \) defined in Section 4.4. The latter is performed by applying the vector model technique for business processes [27]. In this case, it is necessary to define three input entities for the algorithm:

- document collection: two sub-diagram graphs \( G_1, G_2 \in S_k \),
- set of terms \( \Theta \): the union of all graph vertices with their properties listed in Section 3.4,
- weights: for each term \( \theta \in \Theta \) a value of 1 is assigned if the corresponding vertex is a part of the selected subgraph and 0 otherwise.

Considering these two criteria, selecting the appropriate sub-diagram consists either in choosing its group, as shown in Table 3 or specific flow objects that should belong to the component. In the second case, the user is shown a set of similar sub-diagrams, out of which a final selection is needed.

Merge Guidelines Outline

Composition of the BPMN model using the repository of reusable parts should always be conducted with respect to business process modelling guidelines to avoid typical anomalies [28]. Incoherent business process models may be affected by syntactical anomalies which include incorrect usage of activities, gateways, connecting objects or swimlanes. Another group are structural anomalies, related to the improper dynamic behaviour of the process, for example the presence of deadlocks and infinite loops.

Considering the correctness of the composed process model, it is recommended that its generation is supported by a graphical tool connected to the repository. The interface should suggest recommended connection using the position-based classification technique [29]. An example of a forward model completion is presented in Figure 9.

On the other hand, the system should also prevent the user from creating links between flow objects which may lead to structural anomalies. A potential deadlock is shown in Figure 10.

Conclusions

The approach presented in this paper offers a set of activities which support creation of reusable component business process models. The proposed method uses
complete BPMN diagrams as input, based on which the component models are generated.

Such a decomposition algorithm may be used to create a component model repository which facilitates process redesign as well as provides a possibility to avoid common modelling anomalies.

In the future work it is planned to include the message flow in the decomposition which will increase the support of the method for various process elements. Another valuable extension would be to allow generation of sub-diagrams representing a specific set of activities, based on their semantic description.

Fig. 9. BP modelling recommendation based on forward model completion technique. Recommended links were marked with green colour.

Fig. 10. Forbidden connections between sub-diagrams.

References

4. F. Milani et al., Business & Information Systems Engineering 58.1, p.7-17 (Springer, 2016)
22. BPM Initiative Model Collection, http://bpmmai.org/
25. K. Ruohonen, Graph theory, Graafiteoria lecture notes, (TUT, 2013)