Towards Duplication-Free Feature Models when Evolving Software Product Lines

Amal Khtira, Anissa Benlarabi, Bouchra El Asri
IMS Team, SIME Laboratory, ENSIAS, Mohammed V Souissi University
Rabat, Morocco
amalkhtira@gmail.com, a.benlarabi@gmail.com, elasri@ensias.ma

Abstract—Since the emergence of Software Product Line Engineering, the requirements evolution issue has been addressed by many researchers and many approaches have been proposed. However, most studies focused on evolution in domain engineering while application engineering has not received the same attention. During the evolution of a derived product, new features are added or modified in the application model, which may cause many model defects, such as inconsistency and duplication, both in application model and between the latter and the domain model. The aim of this paper is to propose a framework that enables to avoid duplication when evolving software product lines.

Keywords—Software Product Line; Requirements Evolution; Domain Engineering; Application Engineering; Duplication.

I. INTRODUCTION

The Software Product Line Engineering (SPLE) has emerged as a paradigm whose main objective is to develop software applications based on a core platform. The adoption of this approach by companies enables them to reduce time to market, to reduce cost and to produce high quality applications. Another major advantage of the PLE is the reuse of core assets to generate specific applications according to the need of customers.

The SPLE approach consists of two processes, namely, domain engineering and application engineering [1]. During these processes, a number of artefacts are produced which encompass requirements, architecture, components and tests. Domain engineering involves identifying the common and distinct features of all the product line members, creating the design of the system and implementing the reusable components. During application engineering, individual products are derived based on the artefacts of the first process, using some techniques of derivation.

Many issues related to SPLE have been addressed both by researchers and practitioners, such as reusability, product derivation, variability management, etc. The focus of our study will be on SPL evolution. Evolution is defined by Madhavji et al. [2] as “a process of progressive change and cyclic adaptation over time in terms of the attributes, behavioral properties and relational configuration of some material, abstract, natural or artificial entity or system”. This definition applies to different domains, including software engineering.

In the literature, several studies have dealt with evolution in Software Product Lines (SPLs). Xue et al. [3] presented a method to detect changes that occurred to product features in a family of product variants. In order to support agile SPL evolution, Urli et al. [4] introduces the Composite Feature Model (CFM), which consists of creating small Feature Models (FMs) that corresponds each to a precise domain. Other approaches, such Ahmad et al.’s [5], focused on the extraction of architecture knowledge in order to assess the evolutionary capabilities of a system and to estimate the cost of evolution. Some papers focused on the co-evolution of different elements of SPLs [6].

Based on the literature, we have found that most of the studies addressing software evolution focus on domain engineering, while application engineering has not received the same interest. However, the experience has proven in many industrial contexts that systems continue to change even after the product derivation. This change can be the source of many problems in the product line such as inconsistency and duplication. Indeed, the core assets of the product line and the artefacts of derived products are most of the time maintained by different teams. Moreover, developers under time pressure can forget to refer to the domain model before starting to implement the changes. For these reasons and others, duplication in SPL can easily happen. We consider Duplication the fact of adding to the application model features of the same semantics, which means that they satisfy the same functionality. In this paper, we propose a framework that deals specifically with the problem of duplication when evolving products in application engineering.

The remainder of the paper is structured as follows. Section 2 gives an overview of the background of our study and describes the problem we are dealing with. In Section 3, we present the basic concepts and the overview of the proposed framework. In Section 4, we provide a formalization of the basic concepts before describing the algorithm of deduplication. An application of the framework on a case study is presented in Section 5. Section 6 positions our approach with related works. The paper is concluded in Section 7.

II. BACKGROUND

In this section, we introduce the background of our study. First, we present the SPLE paradigm, then we give an insight on the problem of duplication when evolving products in application engineering.

A. Software Product Line Engineering

A SPL is defined by Clements and Northrop [7] as “a set of intensive-software systems sharing a common, managed set of features that satisfy the specific needs of a particular market segment or mission and that are developed from a common set of core assets in a prescribed way”. The main goals of a SPL are to reduce the cost of developing software products, to enhance quality and to promote reusability.

The domain engineering phase of the SPLE framework is responsible for defining the commonality and variability of the applications of the product line. Capturing the common
features of all the applications increases the reusability of the system, and determining the variant features allows the production of a large number of specific applications that satisfy different needs of customers. In order to document and model variability, many approaches have been proposed. Some of them proposed to integrate the variability in the existing models, such as UML models or feature models (FORM [8]). Pohl et al. [1] preferred to define it separately in a dedicated model, i.e., the orthogonal variability model. Another approach proposed by Salinesi et al. [9] used a constraint-based product line language. When the model is ready, the next step consists of creating the design of the system which contains the software components and their relationships. Those components are then implemented and the code of the product line is generated.

The process of creating a specific product based on a SPL is referred to as product derivation or product instantiation. Product derivation consists of taking a snapshot of the product line by binding variability already defined in the domain engineering and using it as a starting point to develop an individual product. This process is applied during application engineering phase and is responsible for instantiating all the artefacts of the product line, i.e., model, design, components, etc.

B. Duplication of Features during SPL Evolution

The goal of SPLE is to make an up-front investment to create the platform. Indeed, during domain engineering, the requirements of all the potential applications are captured, and as far as possible, the scenarios of the possible changes have to be predicted and anticipated. The evolution and maintenance of the product line are conducted through several iterations until the platform becomes as stable as possible. As new evolutions arise, the domain artefacts are adapted and refined.

On the one hand, the team responsible for developing and maintaining the product line studies the requirements of each customer and derives specific applications that respond to these requirements. On the other hand, a different team takes in charge the maintenance of each application. Following the logic of SPLE, the derived applications are not supposed to change much, but the experience has shown that this assumption is not always true. In fact, even after the derivation of a specific product, new demands can be received from the customer, either changes to existing features or addition of new ones.

During the maintenance of a product, duplication of knowledge can easily happen when evolving the model, the design or the code. In [10], four categories of duplication are distinguished:

- Imposed duplication: Developers cannot avoid duplication because the technology or the environment seems to impose it.
- Inadvertent duplication: This type of duplication comes about as a result of mistakes in the design. In this case, the developers are not aware of the duplication.
- Impatient duplication: When the time is pressing and deadlines are looming, developers get impatient and tend to take shortcuts by implementing as quick as possible the requirements of customers. In these conditions, duplication is very likely to happen.
- Inter-developer duplication: Different people working on one product can easily duplicate information.

In the context of SPLE, at least the three last categories might occur. Indeed, when a derived application is shipped, developers responsible for maintaining it do not have a clear visibility of the domain model because another team conceived it. Thus, developers of the application may add features which are already satisfied in the domain model and have only to be derived or configured. In addition, under time pressure, developers do not refer to the application model and might add features which are already implemented. To the best of our knowledge, few attempts have dealt with duplication in the application engineering. The aim and contribution of this paper is to provide a framework that helps developers avoid duplication in a SPL when evolving a specific product.

III. A FRAMEWORK TO AVOID DUPLICATION WHEN EVOLVING DERIVED PRODUCTS

In this section, we first provide a short definition of the basic concepts used in the framework, then we present the overview of the framework.

A. Basic Concepts

Before going any further, we will give an insight of the basic concepts used in the framework.

- Domain Model: A domain is a family of related products, and the domain model is the representation of all the different and common features of these products. There are many types of domain models, but the most interesting are the feature model [8] and the variability model [1].
- Application Model: The model corresponding to an individual application. It is generated by binding the variability of the domain model in a way that satisfies the needs of a specific customer [1].

- Feature: A feature is the abstraction of functional or non-functional requirements that help characterize the system and must be implemented, tested, delivered, and maintained [8][11]. A feature is either:
  - Mandatory: it exists in all products.
  - Optional: it is not present in all products.
  - Alternative (One Of): it specializes more general feature; only one option can be chosen from a set of features.
  - Or: One or more features may be included in the product.

- Variation Point: Variation points are places in a design or implementation that identify the locations at which variation occurs [12].

- Variant: It is a single option of a variation point and is related to the latter using a variability dependency [13].
**Specification**: Requirements specification is a description of the intended behavior of a software product. It contains the details of all the features that have to be implemented during an evolution of the system.

**Specific Variant or Variation point**: We consider a variant or a variation point as specific when they concern a particular need of an application that belongs to the product line (e.g., features related to confidential data, features that need legal authorization).

**Generic Variant or Variation point**: We consider a variant or a variation point as generic if they can be demanded by many applications of the product line (e.g., ergonomic or utility features, non-functional features).

### B. The Framework in a Nutshell

With the large number of features in the SPLs, the manual checking of duplication becomes a complicated and an error-prone task. In order to deal efficiently with the problem of duplication during the evolution of derived products, we propose the framework depicted in Figure 1 as an attempt to set an automated deduplication tool.

![Figure 1. The overview of the framework.](image)

**Initial Specification**: In this framework, we take as an input the specification of a new evolution related to a derived product. This specification contains the requirements that have to be implemented in this specific product. To use these requirements, we need to express them as features using the FODA feature model [14]. In the context of our framework, we consider that a feature is the association of a variation point and a variant. To create the feature model, we opt for FeatureIDE [15]. This tool enables to graphically create the feature model and the associated XML is generated automatically.

**Domain and Application Models**: The main prerequisites of the framework are the domain model and the application model. To create these models, we use the FeatureIDE tool in order to generate the sources in the form of XML files.

**Repository**: The repository contains the features of the domain model and also the set of all the possible synonyms and alternatives for the concepts used in the product line. The elements of the repository are defined using the Resource Description Framework (RDF). RDF [16] is a W3C recommendation that supports semantic interoperability between different resources on the web.

**Deduplication Tool**: This tool contains a set of algorithms of features verification. In this paper, we focus on the algorithm of deduplication. Before describing the algorithm, we need to define some predicates.

**Equivalence**: We consider that a variation point (resp. a variant) is equivalent to another variation point (resp. variant) if they both implement the same functionality, which means that they have the same semantics. We define the function \( \text{Equiv} \) which can take three values:

\[
\text{Equiv}(x) = x_0 \Rightarrow \begin{cases} x_0 = x & x_0 \text{ is equivalent to } x \text{ and } x_0 \neq x \\ x_0 \in \emptyset & \end{cases}
\]

Example: The variant "On-line Sales" associated to the variation point "Sales" is equivalent to the new variant "e-sales" (cf. Section 5).

**Duplication**: We consider that a feature of the specification is duplicated if the associated variation point and variant have equivalents in the application model or the domain model.

The aim of the algorithm is thus to verify the non-duplication of all the features of the initial specification in order to generate a new correct specification. Indeed, for each feature of the initial specification, the algorithm verifies whether the associated variation point and variant have equivalents in the domain model and the application model. The detection of equivalence is carried out based on the Repository content. The steps of the algorithm are explained in details in Section 4.

**Duplication-Free Specification**: The output of the framework is a specification that does not contain features causing duplication in the SPL.

### IV. A Algorithm for Duplication-Free SPL

In this section, we provide the formalization of the basic concepts used in the framework, then we describe the deduplication algorithm.

#### A. Formalizing the Basic Concepts

Prior to explaining the algorithm, a certain number of predicates must be defined. We denote by \( D \) the domain model. PD is the set of variation points of \( D \), and \( VD \) is the set of variants of \( D \).

\[
PD = \{PD_1, PD_2, \ldots, PD_p\} \\
VD = \{VD_1, VD_2, \ldots, VD_q\}
\]

Similarly, we denote by \( A \) the application model of a derived application. PA is the set of variation points of \( A \), and \( VA \) is the set of variants of \( A \).

\[
PA = \{PA_1, PA_2, \ldots, PA_s\} \text{ with } s \leq p \\
VA = \{VA_1, VA_2, \ldots, VA_t\} \text{ with } t \leq q
\]

Thus:

\[
PA \subseteq PD \text{ and } VA \subseteq VD
\]
We denote by $S_0$ the specification of an evolution, i.e., the set of new features to implement.

$$S_0 = \{F_1, F_2, \ldots, F_n\}$$

$P$ and $V$ are, respectively, the sets of variation points and variants, which correspond to the features defined in $S_0$.

$$P = \{P_1, P_2, \ldots, P_k\}, \quad V = \{V_1, V_2, \ldots, V_n\}$$

It has to be noted that $P$ and $V$ are not subsets of $P_A$ and $V_A$. In our framework, we consider that a feature in $S_0$ can be defined as follows: $F_k = (P_i, V_j)$.

**B. The Deduplication Algorithm**

To implement a new evolution, we propose an evolutionary framework that verifies whether a feature of the specification is duplicated in the model and generates in the end a new verified specification. Figure 2 shows the relationship between the specification of an iteration $k$-1, the feature $F_k$ and the resulting specification.

![Figure 2. Relationship between $S_{k-1}$, $F_k$ and $S_k$.](image)

To verify whether a feature $F_k$ is duplicated or not, the algorithm distinguishes six different cases in each iteration (k-1). These cases are represented in Figure 3.

![Figure 3. The figure shows the different cases of variants and variation points: Every pair $(P_i, V_j)$ corresponds to a case i.](image)

**Case 1:** The variation point associated to the feature $F_k$ has an equivalent in $P_A$ and the variant has an equivalent in $V_A$. Consequently, the feature is removed from the specification, but the domain model and the application model do not change.

$$(F_k = (P_i, V_j)) \land (Equiv(P_i) \in P_A) \land (Equiv(V_j) \in V_A) \quad \Rightarrow \quad S_k = S_{k-1} \setminus \{F_k\}$$

**Case 2:** The feature $F_k$ consists of adding a new variant to a variation point that has an equivalent in $P_A$, where an equivalent of the variant exists already in $V_A$. Consequently, the feature must be removed from the specification.

$$(F_k = (P_i, V_j)) \land (Equiv(P_i) \in P_A) \land (Equiv(V_j) \in V_A \setminus VA) \quad \Rightarrow \quad S_k = S_{k-1} \setminus \{F_k\}$$

**Case 3:** The feature $F_k$ requires adding a new variant to a variation point that has an equivalent in $P_A$, and the variant does not have an equivalent in $V_D$. In this case, we assume that $V = V_S \cup V_G$ where $V_S$ is the set of variants of $V$ that are specific to the business of the application, and $V_G$ is the set of variants of $V$ that are generic. We distinguish two sub-cases:

**Case 3.1:** If the variant in question belongs to $V_G$. In this case, the specification does not change and the feature is added directly to the application model.

$$(F_k = (P_i, V_j)) \land (Equiv(P_i) \in P_A) \land (Equiv(V_j) = \emptyset) \land (V_j \in V_S) \quad \Rightarrow \quad S_k = S_{k-1}$$

**Case 3.2:** If the variant in question belongs to $V_G$. In this case, the feature is removed from the specification.

$$(F_k = (P_i, V_j)) \land (Equiv(P_i) \in P_A) \land (Equiv(V_j) = \emptyset) \land (V_j \in V_G) \quad \Rightarrow \quad S_k = S_{k-1} \setminus \{F_k\}$$

The variant $V_j$ is added to the domain model and then to the application model.

**Case 4:** The variation point related to $F_k$ has an equivalent in $P_D$ but not in $P_A$, and the variant has an equivalent in $V_D$ but not in $V_A$. Consequently, the feature must be removed from the specification.

$$(F_k = (P_i, V_j)) \land (Equiv(P_i) \in P_D \setminus P_A) \land (Equiv(V_j) \in V_D \setminus V_A) \quad \Rightarrow \quad S_k = S_{k-1} \setminus \{F_k\}$$

The variation point and the variant must be derived from the domain model and added to the application model.

**Case 5:** The variation point related to $F_k$ has an equivalent in $P_D$ but the variant is new. In this case, the feature is removed from the specification.

$$(F_k = (P_i, V_j)) \land (Equiv(P_i) \in P_D \setminus P_A) \land (Equiv(V_j) = \emptyset) \quad \Rightarrow \quad S_k = S_{k-1} \setminus \{F_k\}$$

The variant is added to the domain model then to the application model.

**Case 6:** The variation point related to $F_k$ does not have an equivalent in $P_D$, and the variant does not have an equivalent in $V_D$. In this case, we assume that $P = P_S \cup P_G$ where $P_S$ is the set of variation points of $P$ that are specific to the business of the application, and $P_G$ is the set of variation points of $P$ that are generic. We distinguish two sub-cases:

**Case 6.1:** If the variation point belongs to $P_S$, the specification does not change and the feature is added directly to the application model.

$$(F_k = (P_i, V_j)) \land (Equiv(P_i) = \emptyset) \land (Equiv(P_i) \in P_S) \quad \Rightarrow \quad S_k = S_{k-1}$$

**Case 6.2:** If the variation point belongs to $P_G$, the feature is removed from the specification, added to the domain model then to the application model.

$$(F_k = (P_i, V_j)) \land (Equiv(P_i) = \emptyset) \land (Equiv(P_i) \in P_G) \quad \Rightarrow \quad S_k = S_{k-1} \setminus \{F_k\}$$
In the end, when all the verifications are carried out for all the features of $S_0$, we obtain $S_N$. This new specification is a duplication-free specification which contains only the features that have to be implemented directly in the application model.

V. CASE STUDY

To illustrate our framework, we propose a part of a feature model of a CRM (Customer Relationship Management). The Figure 4 depicts the domain feature model of the CRM. The feature model is created using FeatureIDE as mentioned in Section 3.

We consider a derived application with the feature model depicted in Figure 5. Based on the XML source, we consider that the tags "and" correspond to variation points and the tags "feature" correspond to variants.

During a new evolution of this application, a number of requirements are demanded by the customer. In our case study, we will take into account only the following requirements:

1) Users can use the application in a disconnected mode.
2) The sector header can contact customers by setting up an appointment.
3) The application must enable users to follow the activity of competitors’ shops.
4) The system must manage e-sales.
5) The sector header can generate summary reports in Excel.

We distinguish the following features:

- $F_1 = (P_1, V_1) = (Connexion, Disconnected mode)$
- $F_2 = (P_2, V_2) = (Contacting Customers, Appointment)$
- $F_3 = (P_3, V_3) = (Shop, Competitor’s store)$
- $F_4 = (P_4, V_4) = (Sales, e-sales)$
- $F_5 = (P_5, V_5) = (Reporting, Excel Report)$

with $S_0 = \{F_1, F_2, F_3, F_4, F_5\}$.

The list of equivalents of the variation points and variants related to these features is described in Table 1.

After applying the verification algorithm to this specification, we came up with the results of Table 2.

*: $V_3$ is specific to this application, because following the activity of competitors’ stores requires a legal authorization, which is not possible for all companies.

**: Generating reports (e.g., in Excel or Word) can be considered as a generic feature, because it can be demanded by other applications.
TABLE II. The Results of the Verification Algorithm

<table>
<thead>
<tr>
<th>Feature</th>
<th>Verification of VP and V</th>
<th>Case</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_1$</td>
<td>$(\text{Equiv}(P_1) \in PA) \land$ $(\text{Equiv}(V_1) \in VD \setminus VA)$</td>
<td>Case 2</td>
<td>$S_1 = S_0 \setminus {F_1}$</td>
</tr>
<tr>
<td>$F_2$</td>
<td>$(\text{Equiv}(P_2) \in PD \setminus PA) \land$ $(\text{Equiv}(V_2) \in VD \setminus VA)$</td>
<td>Case 4</td>
<td>$S_2 = S_1 \setminus {F_2}$</td>
</tr>
<tr>
<td>$F_3$</td>
<td>$(\text{Equiv}(P_3) \in PA) \land$ $(\text{Equiv}(V_3) = \emptyset)$</td>
<td>Case 3.1</td>
<td>$S_3 = S_2$</td>
</tr>
<tr>
<td>$F_4$</td>
<td>$(\text{Equiv}(P_4) \in PA) \land$ $(\text{Equiv}(V_4) \in VA)$</td>
<td>Case 1</td>
<td>$S_4 = S_3 \setminus {F_1}$</td>
</tr>
<tr>
<td>$F_5$</td>
<td>$(\text{Equiv}(P_5) = \emptyset) \land$ $(P_5 \in P_G)$</td>
<td>Case 6.2</td>
<td>$S_5 = S_4 \setminus {F_5}$</td>
</tr>
</tbody>
</table>

In the end, we obtain $S_4 = \{F_3\}$, which means that developers have to implement only the feature $F_3$ in the application model. The other features are all sent to the team maintaining the domain model in order to add the new features and bind the existing ones then re-derive the model.

VI. RELATED WORK

In this section, we provide an overview of the studies most relevant to our work by categorizing them according to the issues addressed in this paper.

Evolution of feature and variability models: In order to reduce complexity and improve the maintainability of variability in large-scale product lines, Dhungana et al. [17] proposes to organize product lines as a set of interrelated model fragments that define the variability of particular parts of the system, and presents a support to semi-automatically merge the different fragments into a complete variability model. The same approach is proposed by Pleuss et al. [18] for feature models. Voelter et al. [19] proposes an approach which consists of separating features in models and composing them by aspect-oriented composition techniques. Cordy et al. [20] defines two particular types of features, regulative features and conservative features, and explains how the addition of these features to the SPL can reduce the overhead of model checking. The common denominator of the cited studies is that they all consider evolution in domain engineering, while our approach deals with evolution in application engineering.

Model Defects in SPL: Several papers in the literature have addressed model defects caused by SPL Evolution. For example, Guo and Wang [21] proposes to limit the consistency maintenance to the part of the feature model that is affected by the requested change instead of the whole feature model. Romero et al. [22] introduces SPLEmma, a generic evolution framework that enables the validation of controlled SPL evolution by following a Model Driven Engineering approach. This study focused on three main challenges: SPL consistency during evolution, the impact on the family of products and SPL heterogeneity. In [23], Mazo provides a classification of different verification criteria of the product line model that he categorizes into four families: expressiveness criteria, consistency criteria, error-prone criteria and redundancy-free criteria. Redundancy can easily be confused with Duplication, but it is completely different, because Mazo focuses on redundancy of dependencies and not redundancy of features. The same study defines also different conformance checking criteria, among which two features should not have the same name in the same model. This is also different from our approach which is based on equivalence and not only equality of features.

Evolution in application engineering: Carbon et al. [24] presents an empirical study which consists of adapting the planning game to the product line context in order to introduce a lightweight feedback process from application to family engineering at Testo, but it does not provide a general approach that is applicable to all SPLs. Hallsteinsen et al. [25] introduces the concept of Dynamic Software Product Lines (DSPL), which provide mechanisms for binding variation points at runtime in order to keep up with fluctuations in user needs. However, this approach does not explain in details how the variability is managed between application and domain engineering. Thao [26] proposes a versioning system to support the evolution of product line and change propagation between core assets and derived products. But this study also does not provide a method to manage features in application engineering. Our approach is different because it provides a feature-oriented approach to manage the evolution of derived products in a way that insures non-duplication in the SPL feature models.

VII. CONCLUSION AND FUTURE WORK

In the literature, many studies have addressed the evolution in SPLs, but the majority of them focused on the domain engineering phase, when application engineering has not been thoroughly discussed. Based on industrial experience, products are also likely to evolve even after their derivation, and this evolution can cause many problems especially duplication in the different artefacts of the product line. In this paper, we provided a framework that deals specifically with duplication in feature models. This framework uses a repository that contains the set of domain features and alternatives of the different concepts of the product line at the aim of verifying the non-duplication of all the features of a new specification. To illustrate the framework, we applied it to a case study from the CRM field. In a future work, we intend to initiate a tool based on the framework architecture, whose objective is to automatize the algorithm of verification and to generate automatically a duplication-free specification that contains only the relevant features to implement.

REFERENCES


