

온톨로지를 이용한 S/W Product line 도메인의 명시적 feature 분석 모델

이순복*, 이태웅**, 김진우*, 백두권*

*고려대학교 컴퓨터학과, **한국과학기술정보연구원

*e-mail:{lsb0510, pkm311, Baik}@software.korea.ac.kr, **minote@kisti.re.kr

Explicit feature analysis model of S/W Product line domain using Ontology

Soon-Bok Lee*, Tae-Woong Lee**, Jin-Woo Kim*, Doo-Kwon Baik*

*Department of Computer Science and Engineering, Korea University,

**Korea Institute of Science and Technology Information

요 약

현재 제품계열 공학에서 feature 중심의 공통성 및 가변성 분석을 통한 재사용성에 대한 연구가 활발히 이루어지고 있다. 지금까지는 도메인 전문가의 직관 및 경험에 의해 feature가 분석되어 그 개념의 불명확함으로 재사용 측면에서 제한점을 내포하고 있다. 본 논문에서는 개별 feature 속성 List 작성을 통해 feature간의 의미관계를 중심으로 한 Pattern 분석 방법을 제시하고, 의미 유사성 관계를 적용한 feature 온톨로지 그래프를 이용하여 S/W 제품계열 도메인 공학에서 사용자와 개발자간의 동일한 해석이 가능하고 재사용성을 위한 명시적 feature를 분석 및 추출하는 모델을 제안한다.

1. 서론

소프트웨어 재사용성(reusability)에 대한 중요성은 제품 개발에 대한 시간 절약 및 비용감소(cost reduction)측면에서 점점 더 중요성이 부각되고 있다. 소프트웨어 공학의 재사용성의 패러다임은 소스 코드(source code)의 재사용에서 컴포넌트 재사용으로 다시 현재는 제품 상위레벨의 도메인 분석의 재사용으로 바뀌고 있다[1]. 하지만, 재사용성측면의 공통성과 가변성 분석 중심의 제품계열공학(Product line Engineering)의 하나인 feature 중심의 소프트웨어 재사용 개발방법론인 FORM[2]은 feature에 대한 분석이 도메인 전문가의 직관에 의한 분석으로 정형화 되어 있지 않아 실제 적용하는 기준이 애매모호하여 많은 feature간의 상호 충돌로 인하여 이해당사자(stakeholder)간의 많은 이해의 오류가 발생하고 있다[4]. 특히 정형화 되어 있지 않는 도메인 분석가의 경험적인 공통성 및 가변성 분석 방법은 규모가 크고 고비용인 소프트웨어를 개발할 시에는 재사용성이 보장 되지 않으며 추후 통합을 위해서 더 많은 비용의 미들웨어를 구축해야하는 비용소모를 내포하고 있다. 따라서 본 논문에서는 feature 속성 List와 feature 온톨로지 그래프를 사용하여 feature 정형화를 통해

feature 분석 시 명시적인 정확성(Correctness)에 입각한 feature 추출 및 stakeholder간의 동일한 해석을 제공하고 feature 간의 중복, 상호 충돌 등의 feature 상호작용 문제(feature interaction problem)를 해결하고자 한다.

2. 관련 연구

2.1 기존의 Feature 도메인 분석 방법의 문제점

FODA에서는 도메인 분석이 feature 중심의 공통점과 차이점을 분석하여 이루어지고 있다[2]. 즉 시스템의 집합 중에서 주도적이고 독특한 feature를 식별하는 것을 바탕으로 도메인을 분석한다. 여기에서 feature는 구현, 테스트되고 배포 및 유지되어야 하는 기능적/비기능적 추상화를 말하는 것으로 요구 사항이나 특징적인 기능을 말한다[3]. 하지만 feature는 일반적인 객체 지향 방법론에 말하는 객체와는 다른 특징적인 것을 가지고 있다. feature는 자신이 지니고 있는 속성에 따라 명확하게 정의 및 분석이 되어져 도메인 분석가들의 공통적인 이해를 통한 명시적인 개념의 feature 추출이 이루어져야 한다. 기존의 feature 분석은 명시적이고, 의미에 기반한 feature 추출 측면에서 한계가 있다.

2.2 Feature분석과 온톨로지의 관계

온톨로지(Ontology)는 “하나의 지식에 대한 공유할 수 있는 이해를 바탕으로 하는 명시적 명세(Formal specification of shared understanding for a knowledge)”를 말한다[6]. 즉 다시 말해서 특정 분야의 지식을 표현하기 위한 기본 지식체계를 말한다. 즉 문서가 내포하고 있는 정보를 동일의미의 용어로 표현한다는 측면에서 “feature”와 유사하고, 특정분야의 지식을 표현하는 체계를 제공한다는 측면에서 “taxonomy”와 유사하며, 특징은 온톨로지로서 정의된 “feature”의 상호 연관관계를 체계적으로 정의 한다는 측면에서 관련성이 있다. 온톨로지는 “concept”와 “relation”로 두 가지로 구성되고, “concept”는 온톨로지 내에 정의된 feature를 말하며, “relation”은 온톨로지 내에 정의된 개념 간의 관계를 정의하는 이원화 관계로 적용된다.

3. 명시적 feature 분석을 위한 온톨로지 적용절차

3.1 적용 프로세스

명시적 feature 추출을 위해 아래 <그림 1>과 같이 총 5단계의 feature 온톨로지 적용 프로세스를 사용해야 한다.



(그림 1) feature 온톨로지 적용 프로세스

단계 1. Usecase 와 Primitive 요구사항간의 비교 매트릭스 정립

도메인 내의 시나리오를 분석하여 최하위 수준의 의미상의 최소 단위인 “semantic primitive 요구사항”으로 나누기 위해 아래 <표 1>와 같이 상호 비교 매트릭스를 활용하여 최초 개발 도메인에 대한 개발자와 사용자간 해당 feature에 대한 의미적 정형성(semantic formalism)을 1차적으로 분석한다.

<표 1>Usecase/primitive 요구사항 간 비교 matrix

Usecase	Usecase1	Usecase2	...	Usecase n
Primitive requirement				
Primitive requirement # 1	feature # 1			
Primitive requirement # 2		feature # 2		
...
Primitive requirement # n				feature # n

단계 2. feature간의 Pattern 분석

1단계의 비교 매트릭스에서 식별된 feature에 대한 의미적 관계를 feature pattern으로 분석한다.

- 규칙 1: 개별 feature 속성 및 가중치 정의

아래 <표 2>와 같이 단계 1에서 식별된 1차 feature의 가중치(weight value)를 부여한 feature 속성 list를 정의한다[6]. 여기서 가중치는 속성들 간의 의미적 관계에 비중을 둔다. <표 2>에서 세부 속성 중 식별형 속성인 이름, 문맥, 핵심어는 의미(semantic)적 측면에서 가장 명확성을 내포하여 가중치의 최대값 1과, 표현형 속성인 문법(syntax) 기반의 feature type, 측정단위는 최소값 0.2를 각각 부여하는 논리적 근거(rationale)를 제시 한다.

<표 2>가중치를 부여한 Meta feature 속성 list

속성 구분	세 부 속 성	가중치(0≤w≤1)
식별형 속성	이름(name)	1.0
	식별자(identifier)	.
	등록기관(registration authority)	.
	동의명(synonymous name)	0.8
정의형 속성	문맥(context)	1.0
	정의(definition)	0.6
관계형 속성	분류체계(classification)	0.8
	핵심어(key word)	1.0
	의존관계(dependency)	0.8
표현형 속성	feature 유형(feature type)	0.2
	측정단위(Unit of measure)	0.2
관리형 속성	책임조직(responsible organization)	.
	등록상태(registration status)	.
	feature 결합시간(feature binding time)	.

- 규칙 2: 개별 feature의 계층적 의미 관계 식별

feature간의 계층적 의미관계는 다음과 같이 4가지 형태로 정의된다.

- 상위/하위 feature간 관계: Is-A/Has-A
- 부분/전체 feature간 관계: Part-of/Has-Part
- 멤버 feature간 관계: Member-of/Has-member
- 구성성분 간 관계: Substance-of/Has-Substance

단계 3. Meta feature 속성 list 작성

1, 2단계에서 식별된 feature 및 각각의 속성 List에 대해서 메타 feature list를 작성한다.

단계 4. Feature 온톨로지 유사성 그래프 사상

생성된 feature들 간의 명시적 의미의 관점에서의 의미 유사도를 측정하기 위하여 feature 온톨로지 유사성 그래프를 이용하며, 의미적 유사성 범위(scope)의 경계조건(semantic boundary)을 중심으로 공통성과 가변성을 분석하는 기준으로 활용된다.

단계 5. 명시적 feature 추출 / 재사용

온톨로지 유사성 그래프를 활용하여 명시적으로 추출된 feature의 속성 List는 표준화된 MFR(Meta Feature Repository)에 저장하여 차기 새로운 feature 추출 시 참조모델(reference model)로 활용한다. 또한 추후 동일/유사 개발 도메인의 S/W 제품 도메인 분석 시 재사용한다.

3.2 feature 온톨로지를 이용한 의미 유사성 매핑

명시적 feature를 추출하기 위한 온톨로지 유사성 그래프는 <그림 2>의 정형적 명세로 정의 된다.

정의 1. (feature 일반적 온톨로지 그래프)

특정 feature 도메인 온톨로지 개념 f에 대해서 상위 개념 U와 하위 개념 d1, d2, ..., dn을 이용하여 온톨로지 feature 그래프를 다음과 같이 U, D, S의 3개의 tuple로 정의된다.

$$Onto(f) := (U, D: \{d_1, d_2, \dots, d_n\}, S) \quad (1)$$

- **U(Upper level ontology)** : feature 온톨로지 상위개념 U이며, 분류적 정의에서 Onto(f)가 상속받는 개념.
 - **D(Down level ontology)** : {d1, d2, ..., dn}는 특정 feature 개념에 대한 하위 개념 집합 D의 원소.
 - **S(Semantics)** : 하나의 feature에 대한 의미적 feature의 집합이며 상위/하위개념 U와 D 사이의 feature 온톨로지 개념의 집합.
- 위 (1)에 대한 선행정의는 아래 집합론적 정형 명세로 정의된다.

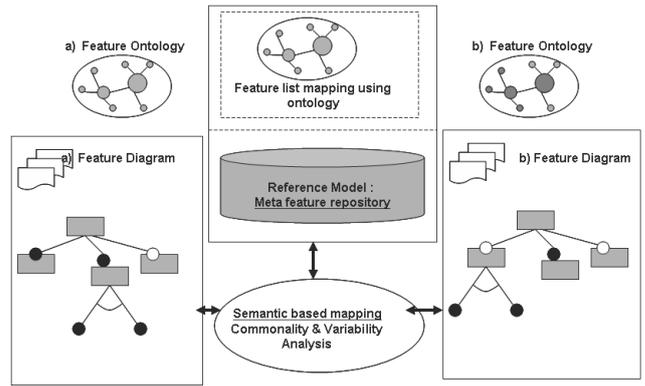
$$f \Leftrightarrow c, S = \left\{ c \mid \bigcup_{f \in MFR} syn(d) \subset syn(f) \subset syn(u), f \in Onto(f) \right\}$$

- **Syn(Synset)** : 2단계 meta feature 리스트에서 정의된 개별 속성 및 관계의 유사적 의미 집합.
- **MFR(Meta Feature Repository)** : 기 식별된 개별 feature리스트에 대한 표준화된 저장고.

정의 2. (feature 온톨로지 유사성 그래프)

Onto(f) := (U, D, S)는 아래 조건을 만족한다.

1. Onto(f)상의 C는 하나의 무한 개념 집합.
(an infinite set of Concepts)
2. $E \subseteq C \times C$ (E : feature 개념간의 edge)
3. $\delta : Feature\ similarity\ Value \rightarrow v$
 $v = \{ \langle c, a, d \rangle \mid \langle c, d \rangle \subseteq C, d \subseteq Sc, \text{ and } a = \delta(Sim\ Val(c, d)) \}$
 위 정의에 기반 하여 feature 온톨로지 그래프는 아래 <그림 2>의 유사성 검색 알고리즘에 의해 사상된다.



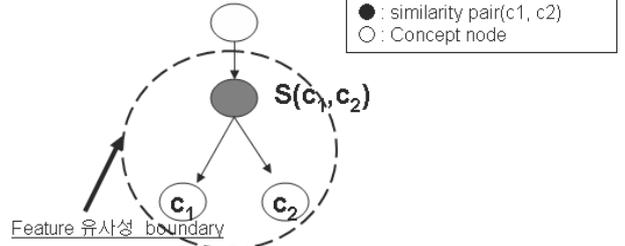
(그림 2) feature 유사성 그래프 이용 의미 사상

정의 3. 명시적 feature 추출 알고리즘

- Child node에 feature 의미 유사성이 있을 경우 : **Child node와 mapping**
- Leaf node에 feature 의미 유사성이 있을 경우 : **Parent node와 mapping**
- leaf node에서 sibling 간 feature 의미 유사성이 있을 경우 : **Sibling node와 mapping**

위 검색 알고리즘은 아래 <그림 3>으로 적용된다[5].

[Feature 유사성 그래프]



```

for ( f=0; f< featureSim(Q_{f1}); f++)do begin
  If match( featureSim(Q_p) ):= true then continue
  else featureConceptMapping(Q_p) do begin
    for all sim(Q_p) ∈ metaFeatureAttribute do begin
      If childMapping() := true then continue
      else if parentMapping() := true then continue
      else if siblingMapping() := true then continue
    end
  end
end
End
return true
    
```

(그림 3) Feature 의미 유사성 검색 알고리즘

유사성 검색 알고리즘에 의해 검색하여 해당 노드의 위치를 찾은 feature의 의미 노드는 아래 공식(2)에 의거하여,

$$Feature\ O(c_1, c_2) = Max_{c \in S(c_1, c_2)} \left[\sum_{n=0}^{\infty} w_v n \right] \quad (2)$$

,where $0 \leq w_v \leq 1$

노드 비교 순서쌍 S(c1,c2) 중 해당 가중치의 합 중 최대값인 feature를 선택하여 의미적으로 명시적인 feature의 산출이 가능하다.

4. 사례 연구

본 논문의 제안된 feature 온톨로지 유사성 그래프의 구체적인 적용 절차를 사례 연구 측면에서 “전자결재 시스템” 도메인에 대하여 적용하였다.

<표 3>는 단계 1에서 개발자와 사용자간 해당 feature에 대한 의미적 정형성(semantic formalism)을 1차적으로 분석한 적용 template이다.

<표 3> Usecase/PR요구사항 간 비교 matrix

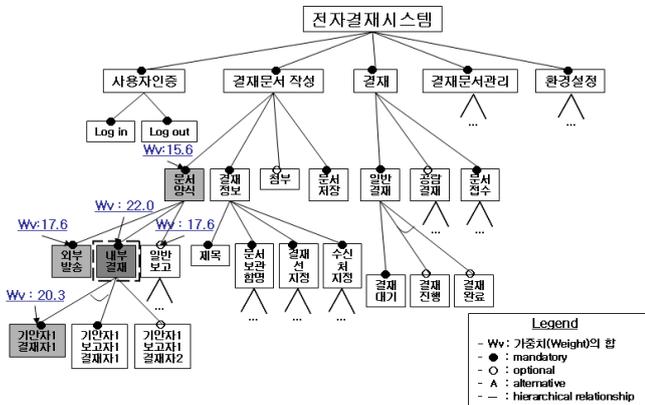
Usecase Primitive requirement	시스템 실행	결재문서 작성	결재 상신	문서 결재	...
...			
결재문서양식의 상황별 템플릿 제공		외부발송 내부결재 일반보고			
대외발송문서 작성 양식 제공		외부발송 선택매뉴			
내부결재문서 작성 양식 제공		내부결재 선택매뉴			
내부결재(기안자,결재자) 양식 제공		기안자, 결재자			
...

<표 4>는 단계 1에서 작성된 비교 Matrix를 이용하여 단계 2의 feature간 Pattern을 분석한다.

<표 4>feature(내부결재) 속성 list의 적용 예제

구분	속 성	적 용	가중치
식별형 속성	이름	내부결재	1.0
	식별자	ISFO01	.
	등록기관	○○정보통신처	.
	동의명	부내결재, 대내결재, 사내결재	0.8
정의형속성	문맥	대내적 결재 문서의 양식 선택	1.0
	정의	대내적인 결재 문서의 작성 양식	0.6
관계형 속성	분류체계	문서양식	0.8
	핵심어	대내, 사내, 부내, 결재	1.0
	의존성	문서양식	0.8
표현형 속성	feature 타입	문자	0.2
	측정단위	Character	0.2
관리형속성	책임조직	○○정보통신과	.
	등록상태	속성이 등록되어 있는 상태 기술	.
	feature 결합시간	시스템 도메인 분석 시 적용	.

분석된 feature Pattern을 표준화된 MFR 과의 feature 온톨로지 유사성 매핑을 수행하여 최적의 명시적 feature를 추출한다.



(그림 4) 가중치를 적용한 온톨로지 유사성 매핑

위 <그림 4> 가중치를 적용한 온톨로지 사상으로 명시적인 feature를 추출하는 과정을 보여준다. 추출의 논리적 근거는 표 4에서의 적용 예제 feature인 “내부결재”의 가중치의 합은, 그 상위 feature (Parent node) 문서양식의 가중치 합이 최상위 feature(Root node)로부터 15.6의 값을 가지므로 표 5에서의 최대 가중치 6.4를 합하여 22.0을 가지고, Sibling node인 외부발송, 일반보고의 가중치 합은 17.6을 각각 가지므로, 명시적인 feature를 추출할 수가 있다.

5. 결론

본 논문에서는 S/W 제품개발 도메인 개발 시 적용되는 feature 분석 방법이 도메인 전문가의 직관적인 분석으로 비정형적임에 따라 실제 적용하기 불명확하고 feature 상호간의 많은 충돌로 이해당사자간의 이해의 오류가 발생하고 있는 이러한 문제를 해결하고자 feature 속성 List 작성과 feature 유사성 온톨로지 그래프를 이용하여 명시적인 feature를 추출하기 위한 모델을 제시하였다[3][4]. 본 논문의 의의는 정확성(Correctness)에 입각한 명시적 feature 추출로 사용자, 개발자간의 feature에 대한 동일한 해석 제공 및 feature 상호 작용간의 문제를 해결하고 재사용성을 증가시키는 데 있다. 향후 연구로는 OWL을 이용한 feature 온톨로지 자동 생성 툴을 적용하여 모델을 검증하고 확장하여 계층적 feature 온톨로지 분석모델을 개발하고자 한다.

참고문헌

[1] Clements, P. and Northrop, L., “Software Product Lines: Practices and Patterns”, Addison-Wesley, Upper Saddle River, 2002.
 [2] Kang, K. C., et al., “FORM: A Feature-Oriented Reuse Method with Domain Specific Reference Architectures”, Annals of Software Engineering, Vol. 5, pp. 143-168.
 [3] Lee, K., K. C., et al., “Concepts and Guide-lines of Feature Modeling for PLSE”, Springer-verlag, pp. 62-77, 2002.
 [4] 송재승 외 2명, “Product line 개발에서 feature 모델의 명세화 기법”, 정보과학회 논문지 제30권 제10호, pp. 1001-1014, 2003. 8.
 [5] 하상범 외 1명, “온톨로지 기반 추론을 이용한 시멘틱 검색 시스템”, 정보과학회 논문지 제32권 제3호, pp. 202-213, 2004. 12.
 [6] Ju-hum Kwon, Doo-Kwon Baik, et al., “Measuring Semantic Similarity Based on Weighting attribute of Edge Counting”, LNCS 3397, 2005.