

Improvements to DSM Clustering Approach for PBS Construction

김찬목

아주대학교 시스템공학과
kcmtg@nate.com

박영원

아주대학교 시스템공학과
ywpark@ajou.ac.kr

Abstract: A top-down scenario based approach is a typical method generating the product breakdown structure (PBS) in systems engineering. While it takes time and effort to perform the scenario based PBS construction, it is the safest approach not to omit any functions. On the other hand, a design structure matrix (DSM) approach to PBS construction is a bottom-up approach based on modularization concept that provides a merit of quick PBS construction while the sanity of its results may not be assured.

Accordingly, this study presents trade-off results between the two approaches to uncover problems in the DSM-based clustering method and proposes improvement approaches to enable the interfaces with external systems, static elements, as well as a new capability to handle the large number of clustering elements.

서론

PBS는 시스템을 구성하는 구성품들을 계층적 구조로 표현한 것으로서 시스템 설계의 결과물이다. PBS를 구축하는 방법에는 시스템 엔지니어링에서 제시하는 시나리오 기반 PBS방법론이 있다. 시나리오 기반 PBS 구축 방법론은 시스템이 어떻게 사용될지를 시간의 흐름에 따라 기술해 봄으로써 시스템이 갖추어야 할 기능이 무엇인지를 먼저 찾아내는 방법으로써, 시스템 수준에서 하부 구성품 수준으로 분해되어 가는 Top-down 방식이다. 이 방법론은 특히 새로운 시스템을 개발할 때 아주 유용한 방법이다. 그러나 시나리오 기반 PBS 구축 방법론은 많은 시간과 노력을 요구한다. 각 수준별로 시나리오를 작성하게 되고 또한 시나리오 작성에서 드러난 기능들을 어떻게 물리적 구성품으로 구현할지에 대해서는 또 다시 여러 대안들을 비교 분석하여 결정하게 된다. 따라서, 시나리오에 기반한 PBS 구축 방법론은 요구사항에 기반하여 누락되는 기능이 없다는 장점을 가지는 대신에 많은 시간과 노력을 요구하게 된다.

이에 비해 DSM Clustering 분석에 기반한 PBS 구축 방법론은 기존의 시나리오 기반 PBS 구축 방법론에 비해 효과적이고 빠르게 구축할 수 있다. PBS를 구축하는데 있어서 DSM 방법론은 먼저 적정 수준의 구성품들을 먼저 선정하게 된다. 적정 수준의 구성품들이 선정되면 선정된 구성품들 간의 인터페이스를 식별한 후, 이를 모듈화 개념에 입각하여 밀접한 연관 관계를 가지는

구성품들을 그룹화시켜 나가는 Bottom-up 방식이다. 여기서 구성품들 간의 연관 관계를 비교하여 연관 관계가 높은 구성품들을 그룹화하는 행위가 Clustering이며, 이러한 Clustering은 Clustering 알고리즘에 따라 컴퓨터를 활용하여 빠르게 수행된다. 따라서 DSM 기반 Clustering 방법론에 의한 PBS 구축 방법은 컴퓨터를 활용한 Clustering 분석을 이용하여 시나리오 기반 PBS 구축 방법론에 비하여 훨씬 빠르게 적절한 PBS 결과물을 얻을 수 있다.

그러나 DSM 기반 Clustering 방법론은 Clustering 알고리즘에 이용하여 빠르게 PBS를 구축할 수 있다는 장점을 가지지만 시나리오 기반 PBS 구축 방법론과 같이 오랜 시간에 걸쳐 검증되지 않은 관계로 몇몇 문제점을 가지고 있다. 따라서 본 논문에서는 동일한 철도차량 시스템을 대상으로 하여 시나리오 기반 PBS 구축 방법론에 의해 생성된 PBS 결과와 DSM 기반 Clustering 방법론에 의해 구축된 PBS 결과를 비교 분석함으로써, DSM 기반 Clustering 방법론의 문제점을 도출하고 이를 보완하는 방법에 관해 다룬다.

또한 DSM 기반 Clustering 방법론을 이용한 PBS 구축 결과는 Clustering 알고리즘에 많은 영향을 받기 때문에 DSM 기반 Clustering 방법론을 개선하는 연구는 기존의 Clustering 알고리즘 성능을 개선할 수 있는 연구가 수반되어야 하며, 이에 본 논문에서는 기존의 Clustering 알고리즘들을 분석하여 보다 효율적인 DSM 기반 Clustering 알고리즘을 제안하고자 한다.

DSM 기반 Clustering 방법론의 문제점 및 개선

이 절에서는 동일한 철도차량 시스템을 대상하여, 시나리오 기반 시스템 아키텍팅 방법론과 DSM 기반 Clustering 방법론으로 각 방법론의 절차에 따라 PBS를 구축하고, PBS 결과의 차이점을 분석하여 PBS 구축에 활용되는 DSM 기반 Clustering 방법론의 문제점을 도출하고 이를 개선하는 부분에 대해서 다룬다.

시나리오 기반 시스템 아키텍팅 방법론에 의한 PBS 구축. 시스템 엔지니어링에 의한 시스템 설계는 요구사항 분석으로부터 시작된다. Martin은 설계 프로세스를 그림 1과 같이 정의하고 있다. 시스템 설계의 시작은 이해관계자의 요구사항을 수집하는 것으로부터 시작되며, 이를 이용하여 시스템의 예상되는 시나리오를 작성하게 된다. 시나리오 작성 이유는 시간의 흐름에 따라 시스템의 거동과 사용자의 거동을 예상함으로써 빠짐없이 시스템의 기능을 도출하기 위해서이다. 이 때 수집된 요구사항과 시나리오를 기반으로 하여 도출된 기능은 동적 기능 및 정적 기능을 모두 포함한다.

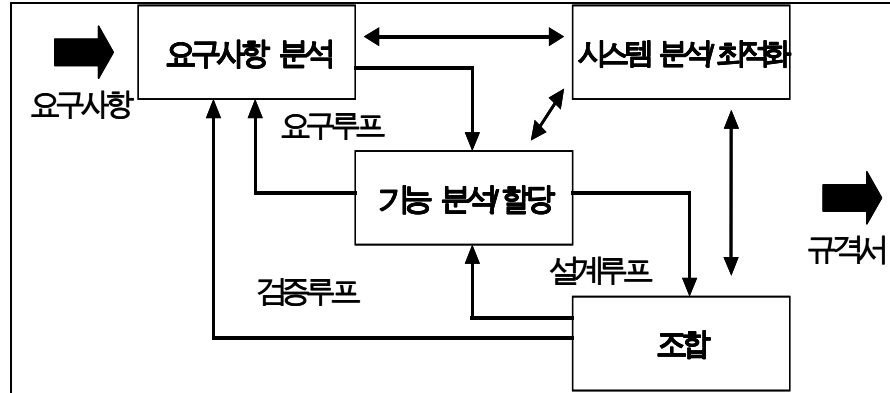


Fig 1 System design process

Martin은 기능(Function)을 “원하는 결과(outcome)를 얻기 위해서 수행되는 과업(task) 및 활동(activity), 또는 시스템 및 시스템 요소의 능력(capability)”으로 정의하고 있으며, 이 기능은 정적(static)이거나 동적(dynamic)일 수 있다고 설명하고 있다. 여기서 동적 기능이라 함은 어떤 입력물을 받아서 가공/처리/변환을 통해 출력물을 제공하는 기능을 말하고, 정적 기능이라 함은 가공/처리/변환 과정이 없는 기능을 말한다. 이렇게 도출된 정적 및 동적 기능은 시스템 분석 및 최적화를 수행하여 조합 단계에서 하부 시스템이 결정되게 된다.

이와 같이 시스템 엔지니어링의 시나리오에 기반한 PBS 구축 방법은 요구사항 분석/기능분석/조합 활동의 반복적인 루프에 의해 시스템 수준에서 하부 시스템 수준으로 분해되는 “Top-down” 방식이다. 철도차량 시스템을 대상으로 PBS 구축을 수행하면 제일 먼저 시스템의 경계를 정의하고 시스템과 외부 시스템을 정의하게 된다. 이에 따라 철도차량 시스템의 시스템 정황도(System Context Diagram)를 그려 보면 아래 그림 2와 같다.

그림 2와 같이 철도차량 시스템의 외부 시스템은 운전자, 승객, 전력 시스템, 신호/통신 시스템으로 정의될 수 있다. 시스템과 외부 시스템이 정의되면 시스템 시나리오를 작성한다. 그림 3은 승객 및 철도차량 시스템에 대한 시스템 시나리오를 FFBD(Function Flow Block Diagram) 형식으로 표현한 것이다.

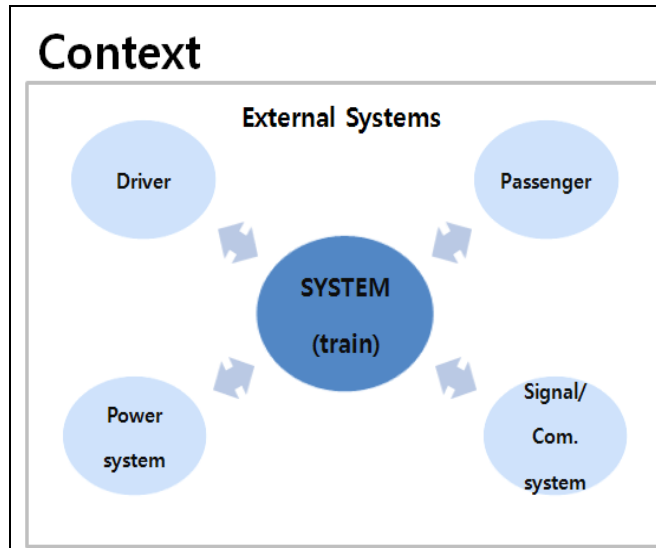


Fig 2 The context diagram of railroad car system

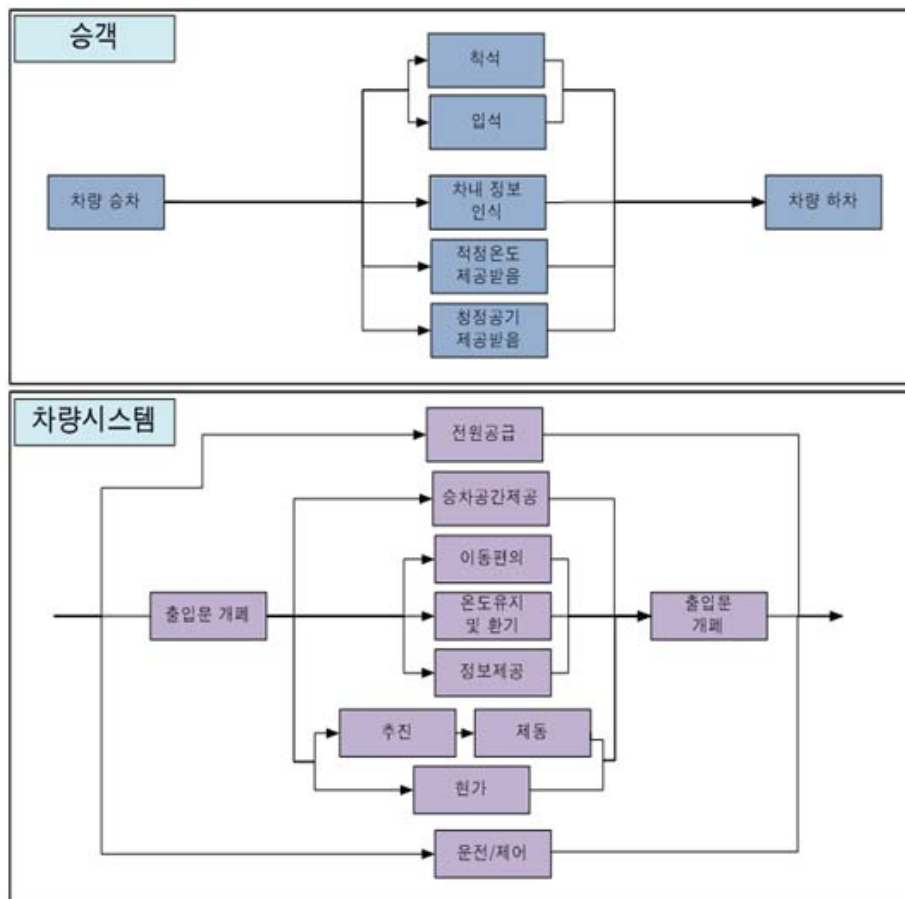


Fig 3 The system scenario of railroad car system

그림 3과 같은 시나리오를 바탕으로 시스템의 기능을 정의하고 조합 단계를 거쳐서 하부 시스템을 정의한다. 이러한 반복 작업에 의해 철도차량 PBS 구축을 수행하면 표 1과 같은 PBS를 구할 수 있다. 표에서 보듯이, 시스템 수준의 설계 프로세스를 통해 정의된 5개의 하부 시스템(level 1)에 대해서 각각의 하부 구성품(level 2)들이 정의되어 있다.

Level 0	Level 1	Level 2
철도차량 시스템	운전 및 제어부	종합제어장치
		자동열차제어장치
		자동열차운전장치
		운전실설비
	객실부	차체
		창문
		출입문
		내장
		차량간 통로연결막
	전원부	집전기
		보조전원
	편의부	의자
		객실손잡이
		승객안내장치
		냉난방 및 환기장치
	주행부	대차틀
		주공기 압축기
		연결부
		견인장치
		현가장치
		추진장치
		제동장치
		동력전달장치

Table 1 The PBS of railroad car system(scenario based)

DSM 기반 Clustering 방법론에 의한 PBS 구축. DSM 기반 Clustering 방법론이 시나리오 기반 시스템 아키텍팅 방법론과 가장 큰 차이를 보이는 것은 “Bottom-up” 방식으로 진행된다는 것이다. DSM방법론을 적용하여 PBS 구축을 수행할 때 가장 먼저 수행하는 일은 적정 수준의 구성품을 선택하는 것이다. 철도차량 시스템의 PBS를 구축하기 위해서 표 2와 같이 총 30개의 구성품을 선택하였다.

번호	구성품 명칭	번호	구성품 명칭
1	종합제어장치	16	의자
2	자동열차제어장치	17	객실 손잡이
3	자동열차운전장치	18	정보 안내 장치
4	운전실 설비	19	HVAC
5	언더프레임	20	대차 프레임
6	사이드 프레임	21	견인장치
7	지붕 프레임	22	공기압축기
8	끝단체	23	연결장치
9	창문	24	현가장치
10	출입문	25	인버터
11	내장	26	추진모터
12	차량간 통로연결막	27	브레이크
13	집전기	28	휠
14	충전기	29	동력전달장치
15	배터리	30	윤축

Table 2 Target components for DSM clustering

구성품이 선택되면 이를 DSM(Component-based DSM)에 표기한다. Component-based DSM에서 구성품 간의 상관 관계는 크게 4가지(에너지 교환, 정보 교환, 물질 교환, 물리적 인터페이스)로 구분된다. 4가지 상관 관계에 대해서 구성품들 간의 상관 관계의 강약에 따라 0점(상관 관계 없음), 5점(상관 관계가 다소 있음), 10점(상관 관계가 강함)을 부여한다. 상관 관계

점수가 아래 그림 4와 같이 부여되면, 이 상관 관계 점수를 대상으로 Clustering 분석을 수행한다. 아래 그림 5는 그림 4에 대한 Clustering 분석을 수행하여 서로 연관성이 높은(강한 상관 관계를 가지는) 구성품들을 묶어서 상부 구성품(Cluster)을 결정한 모습이다.

철도차량 시스템에 대한 DSM Clustering 분석을 수행하여 얻은 Cluster를 바탕으로 하여 철도차량 시스템의 PBS를 표 형식으로 정리하면 아래 표 3과 같다. 2번 Cluster의 경우 내부의 모든 구성품들의 기능을 대변하는 상부 기능을 찾을 수가 없다.

	제어	회전	자동	모터	인버터	사이드	지붕	끝단	창문	인양	내장	자랑	집전기	충전기	배터리	의자	객실	정보	HVAC	대자	건인	공기	연결	현가	인버터	추진	브레이크	축	동력	운축
종합제어장치	10	10	5						5			5	5	5				5	5			5	5	10	5	5				
자동열차제어장치	10	10	5																											5
자동열차운전장치	10	10	5						5																					5
운전실 열비	10	10	5																											5
언더프레임					10	10	10																							
사이드 프레임					10	10	10	5	5	5																				
지붕 프레임					10	10	10													5										
끝단					10	10	10				5																			
창문					5																									
출입문		5	10	5	5																	10								5
내장					5														5											
차량간 통로연결막						5																								
집전기				5																										5
충전기	5															10														10
배터리														20																
의자										5																				
객실 손잡이										5																				
정보 안내 장치			5	5						5																				5
HVAC	5			5		5				5																				5
대자 프레임																						10	10	10	10			10	10	10
건인장치																						10								
공기압축기	5																					10								
연결장치																						10								
현가장치																						10								
인버터	5	10	10										10																	20
추진모터																						10								20
브레이크	5	10																				10								10
축																														10
동력전달장치																														10
운축																														20

Fig 4 The DSM of railroad car system

IV. 클러스터링(clustering) 결과																																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30		
언더프레임	10	10	10																													
사이드 프레임	10	10	10	5	5										5																	
지붕 프레임	10	10	10							5																						
끝단체	10	10	10				5																									
창문		5																														
내장		5								5																						
차량간 통로연결막				5																												
의자						5																										
객실 손잡이						5																										
HVAC			5			5					5				5					5												
종합제어장치										5	10	10	5	5	5	5	10	5			5		5	5	5	5	5					
자동열차제어장치											10	10	5					5														
자동열차운전장치											10	10	5	5	5			5														
운전실 설비											10	10	5				5															
출입문		5									5	10	5				5					10										
집전기													5				5															
정보 안내 장치					5							5	5					5														
인버터										5	10	10			10			20														
추진모터																		20	10													
대차 프레임																			10	10	10	10	10	10	10					10		
견인장치																			10	10	10	10	10	10	10							
공기압축기										5									10	10	10	10	10	10	10							
연결장치																			10	10	10	10	10	10	10							
현가장치																			10	10	10	10	10	10	10							
브레이크										5	10								10	10	10	10	10	10	10							
충전기										5									10													
배터리																															20	
휠																															10	10
동력전달장치																															10	10
윤축																															20	20

Fig 5 The result of clustering for railroad car components

Level 0	Level 1	Level 2	Level 3
철도차량 시스템	Cluster 1 (객실부)	차체	언더프레임
			측면구조들
			지붕구조들
			끝단막이구조들
		창문	
		출입문	
		내장	
		차량간 통로연결막	
		의자	
	객실 손잡이		
	HVAC		
	Cluster 2	운전 및 제어부	종합제어장치
			자동열차제어장치
			자동열차운전장치
			운전실설비
		출입문	
		집전기	
		정보 안내 장치	
	추진부	인버터	
추진모터			

	Cluster 3 (전원부)	충전기	
		배터리	
	Cluster 4 (동력전달부)	휠	
		동력전달장치	
		윤축	
	Cluster 5 (주행부)	대차 프레임	
		견인 장치	
		공기 압축기	
		연결장치	
		현가장치	
			브레이크

Table 3 The railroad car's PBS resulted from DSM clustering

DSM 기반 Clustering 방법론의 문제점 및 개선. 시나리오 기반 시스템 아키텍팅 방법론에 의해 도출된 철도차량 시스템의 PBS 와 DSM Clustering 분석에 의해 도출된 철도차량 시스템 PBS 의 가장 큰 차이점은 “편의부”에서 찾을 수 있다.

첫 번째로, DSM Clustering 분석에 의해 도출된 철도차량 시스템의 PBS 를 보면 “정보 안내 장치”와 “HVAC”이 “편의부”로 따로 Cluster 를 형성하지 못하고 Cluster 1(객실부)에 포함되어 있다.

이렇게 시나리오 기반 시스템 아키텍팅 방법론의 결과와 차이를 보이는 이유는 기존의 DSM 기반 Clustering 방법론의 경우 Clustering 분석 대상이 되는 DSM 요소를 오직 대상 시스템으로 한정하고 있기 때문이다.

“정보 안내 장치”는 승객에게 현재 역 및 다음 역에 대한 정보를 제공하고, “HVAC”은 승객에게 적정 온도를 유지하는 공기와 청정 공기를 끊임없이 제공한다. 그러나 기존의 DSM 방법론의 경우에 DSM 요소를 Clustering 분석 대상 시스템만으로 한정하고 있기 때문에 승객과 정보 안내 장치와의 정보(information) 교환과 승객과 HVAC 간의 물질(material) 교환을 표기할 수가 없다. 따라서 승객과 HVAC, 정보 안내 장치가 하나의 Cluster 로 묶일 수 있는 기회 자체가 없는 것이다.

두 번째로, DSM Clustering 분석에 의해 도출된 철도차량 시스템의 PBS 를 보면 “의자”와 “객실 손잡이”도 역시 “편의부”로 따로 Cluster 를 형성하지 못하고 Cluster 1(객실부)에 포함되어 있다.

이러한 차이를 보이는 이유는 앞서 설명한 바와 같이 기존의 DSM 기반 Clustering 방법론에서는 외부 시스템과의 연관 관계를 고려하지 않는다는 문제점과 함께, 기존의 4 가지 상관 관계(에너지 교환, 정보 교환, 물질 교환, 물리적 인터페이스)로는 정적 기능을 수행하는

구성품들 간의 상관 관계를 표현할 수 없다는 문제점 때문에 발생한다.

“의자”와 “객실 손잡이”는 모두 탑승한 승객이 편리하고 안전하게 목적지까지 이동하는데 편의를 제공하는 기능을 수행하고 있다. 의자와 객실 손잡이는 어떤 입력물을 처리/변환하여 출력물을 전달하지는 않는다. 그러므로 기존의 관계 카테고리에 따르면 의자와 객실 손잡이는 승객과 어떠한 에너지 교환이나 정보 교환, 물질 교환이 없으며, 관계 표시가 없기 때문에 승객과 의자, 객실 손잡이는 하나의 Cluster로 묶일 수가 없다. 그러나 의자와 객실 손잡이는 어떤 입력물을 처리/변환하여 출력물을 전달하지는 않지만 승객에게 편의를 제공하고 있음은 틀림없다. 그러므로 의자, 객실 손잡이와 같이 정적 기능을 수행하는 구성품이 승객과 하나의 Cluster를 이루기 위해서는 새로운 관계 카테고리가 필요하다.

위에서 기술한 PBS 구축에 활용되는 기존의 DSM 기반 Clustering 방법론의 문제점들을 개선하기 위해서 다음과 같이 2가지를 제안한다.

1. DSM 기반 Clustering 방법론을 이용하여 PBS를 구축할 경우, Clustering 대상 요소로 외부 시스템을 포함시켜 Clustering을 수행한다.

Clustering 대상 구성품들 중에는 대상 구성품들 간의 상관 관계보다 외부 시스템과의 상관 관계가 중요할 수가 있다. 그러므로 Clustering 분석 대상 구성품에 외부 시스템을 통합한 후 Clustering 분석을 수행한다. Clustering 분석 결과로 나온 Cluster는 외부 시스템을 포함할 수도 있다. 이럴 경우에 Cluster의 요소 중 외부 시스템을 제외한 나머지를 진정한 Cluster 요소로 간주하면 된다. 예를 들어서, 철도차량 시스템 구성품들과 외부 시스템을 통합하여 Clustering을 분석한 결과 Cluster 1이 승객, 정보 안내 장치, HVAC로 구성되었다면, 승객을 제외한 정보 안내 장치와 HVAC만을 Cluster 1의 요소로 간주한다는 것이다.

2. 정적 기능을 수행하는 구성품의 관계를 표시할 수 있는 “서비스(service)”라는 새로운 관계 카테고리를 생성한다.

정적 기능을 가지는 구성품은 입력물을 변환/처리/가공하여 뚜렷한 출력물을 제공하지 않지만 이들도 어떤 목적을 가지고 있으며, 목적 수행을 위해 어떤 요소와 관계를 가진다. 그러므로 정적 기능을 가지는 구성품들의 상관 관계를 표시하기 위해서 “서비스(service)”라는 새로운 관계 카테고리를 생성한다. 이러한 정적 기능을 수행하는 구성품들은 대부분 사용자의 안전을 확보하거나 편의를 제공하는 기능들을 수행하므로 서비스란 관계 카테고리 로 대부분 포함시킬 수 있을 것이다.

DSM 기반 Clustering 알고리즘의 문제점 및 개선

PBS 구축에 적용되는 DSM 방법론의 핵심은 Clustering 분석에 있다. Clustering 분석은 수많은 반복 작업을 통해 각각의 결과값 중 가장 좋은 결과값을 최종 결과값으로 선택하는 분석법이다. 따라서 Clustering 분석은 사람의 손으로는 계산하기 어려우며, 이러한 이유로 인해 모든 Clustering 분석은 컴퓨터를 이용한 자동 프로그램에 의해 수행된다. 요약하자면 PBS 구축에 적용되는 DSM 방법론의 핵심은 Clustering 분석에 있고, Clustering 분석은 자동 프로그램에 의해 수행되므로 Clustering 분석은 결국 자동화 프로그램의 Clustering 알고리즘에 달려 있다는 것이다.

이에 따라 본 단락에서는 기존에 연구된 Clustering 분석 프로그램의 알고리즘을 분석하여 문제점을 도출하고, 이를 개선하는 알고리즘을 제안하고자 한다.

PBS 구축에 활용되는 Clustering 알고리즘의 특징. PBS 구축에 이용되는 Clustering 문제의 특징을 정리하면 다음과 같다.

1. 두 객체 간의 상관 관계만을 제공 받는다. 생산시스템의 기계들을 셀로 묶는 그룹 테크놀러지와 같이 각 객체들을 특성 벡터로 표현할 수 없으며, 다만 두 객체 간의 상관 관계만을 제공받는다.
2. 하나의 객체는 하나의 Cluster에만 속한다. PBS에서 하나의 구성품은 하나의 상부 구성품(또는 하부 시스템)에만 속하며 여러 상부 구성품에 동시에 속할 수 없다.
3. Cluster 내 상관 관계와 Cluster 간 상관 관계를 동시에 고려해야 한다. 이는 임동순의 구분 방법에서 $\lambda=0.5$ 인 문제와 동일하다.
4. 사전에 Cluster 개수를 정하지 못한다. 파라미터 접근법의 Clustering 알고리즘의 경우 대부분의 알고리즘이 초기에 Cluster 개수를 정하고 시작하게 된다. 몇몇 Cluster의 개수를 알지 못하는 상태에서 시작하는 알고리즘이 있으나 이는 PBS 구축에 이용되는 Clustering 문제에는 적용하기가 다소 어렵다. 이 내용은 다음 단락에서 살펴 보기로 하겠다.
5. 최적값을 구하기 보다는 근사 최적값을 구한다. Clustering 문제에서 해가 될 수 있는 경우의 수는 아래 수식 1과 같다. 아래 수식에서 보듯이 Cluster의 개수를 모를 경우에는 대안의 수가 더욱 증가하게 되며, 이로 인해 최적값을 구하기 보다는 한정된 시간에 근사 최적값을 구하는 연구로 많이 진행되고 있다.

$$\sum_{M=1}^N \left(\frac{1}{M!} \sum_{j=1}^M (-1)^{M-j} {}_M C_j \times j^N \right)$$

- M: Cluster 개수
- N: 전체 객체 수
- ${}_M C_j$: 조합

Cluster 개수를 알 경우, 해의 대안 수

Cluster 개수를 모를 경우, 해의 대안 수

(1)

PBS 구축에 활용되는 기존 Clustering 알고리즘의 종류. PBS 구축을 위한 Clustering 문제와 같이 객체 간의 상관 관계만을 알 수 있으며 초기에 Cluster 개수를 알지 못하는 경우에는, 일반적으로 Cluster 의 개수를 증가시켜 나가면서 각 단계마다 부분 최적화를 수행하는 반복적 실험에 의한 Clustering 방법들을 사용한다. 반복적 실험에 의한 Clustering 알고리즘의 프로세스는 아래 그림 6 과 같다. 아래 그림 6 과 같이 Cluster 의 개수를 서서히 증가시켜 나가면 목적함수 값이 서서히 증가(또는 목적함수에 따라 감소)하다가 어느 수준에 이르면 다시 감소(또는 증가)하게 되며, 이 시점의 Cluster 개수를 최적의 Cluster 개수로 결정하게 된다. 따라서 이러한 반복적 실험에 의한 Clustering 방법은 반복 실험에 의한 시간 비용이 커진다는 단점을 가진다.

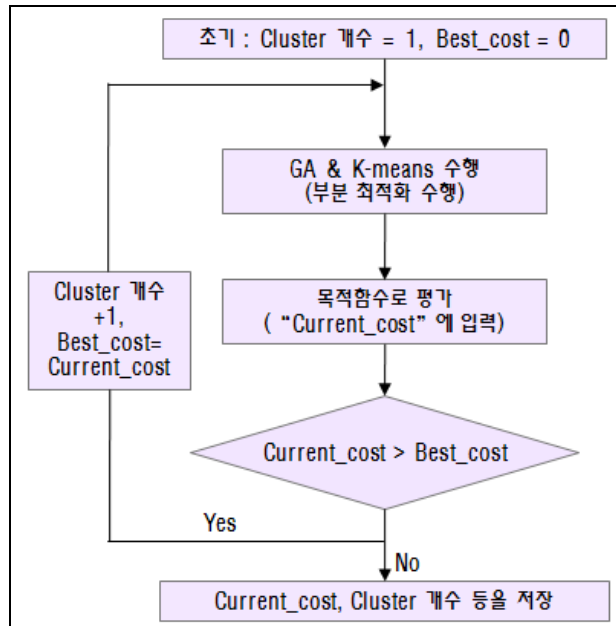


Fig 6 Clustering algorithm by using a repetitive test

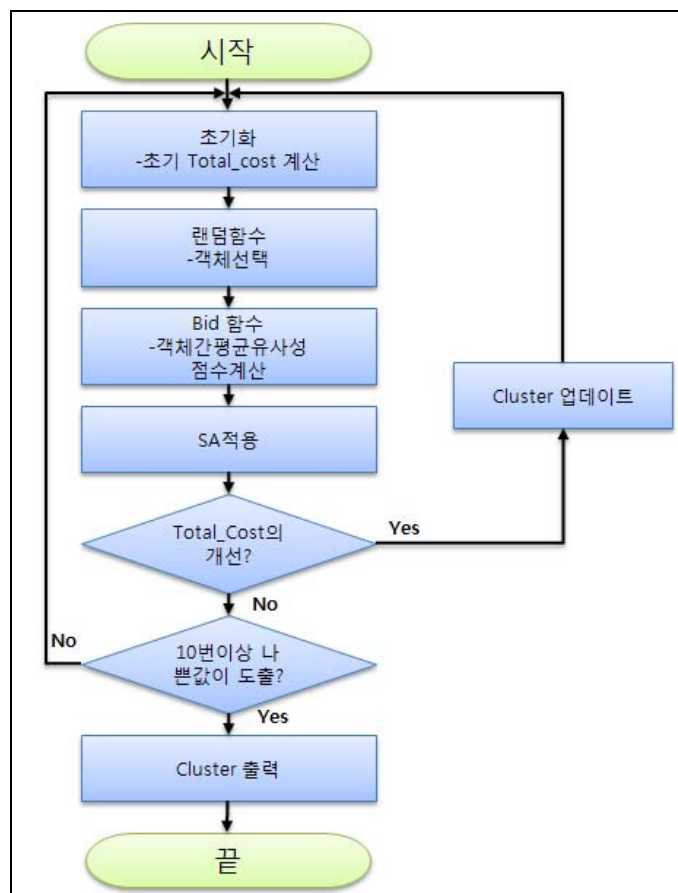


Fig 7 Flowchart of the Thebeau's clustering algorithm

Ronnie E. Thebeau 는 Fernandez 의 알고리즘을 개선하여 DSM 기반 Clustering 알고리즘을 고안하였다. Thebeau 알고리즘의 특징은 랜덤 함수를 사용하여 Clustering 대상 객체를 무작위로 선택한다는 것이다. 이러한 이유는 Clustering 객체를 선택함에 있어서 어떤 규칙을 정하게 되면 최적값을 구하지 못하는 경우가 발생하기 때문이다. 이러한 무작위 객체 추출법을 사용함으로써 근사 최적값이 아닌 최적값을 구할 수 있다는 장점을 가지지만 동시에 Clustering 대상 객체가 많아질수록 최적값을 얻기 위한 반복 계산시간이 길어진다는 단점을 가진다. Thebeau 알고리즘의 프로세스는 위 그림 7 과 같다.

기존 Clustering 알고리즘의 문제점 및 개선. PBS 구축을 위해 사용되는 반복적 실험에 의한 Clustering 알고리즘은 반복 실험에 의한 시간 비용이 커진다는 단점을 가진다. 반면 Thebeau 알고리즘은 랜덤함수를 이용한 무작위 객체 추출법으로 최적값을 구할 수 있다는 장점을 가지지만 역시 Clustering 대상 객체가 커지면 좋은 해를 얻기 위해서 수많은 반복 연산을 수행해야 한다는 단점을 가진다. 즉 Thebeau 알고리즘도 역시 Clustering 대상 객체가 커지면 몇 회의 반복 연산으로 최적값을 도출할 수 있는지 알 수 없으므로 결국 정해진 반복 연산 결과 중 최소값을 선택할 수 밖에 없다.

따라서, 본 논문에서는 Clustering 대상 객체의 수가 많아질 경우에도 좋은 품질의 근사 최적값을 빠른 시간에 도출할 수 있는 알고리즘을 고안하는 것을 목표로 한다. Thebeau 알고리즘과 제안하는 알고리즘의 성능 차이를 그림으로 표시하면 아래 그림 8 과 같다.

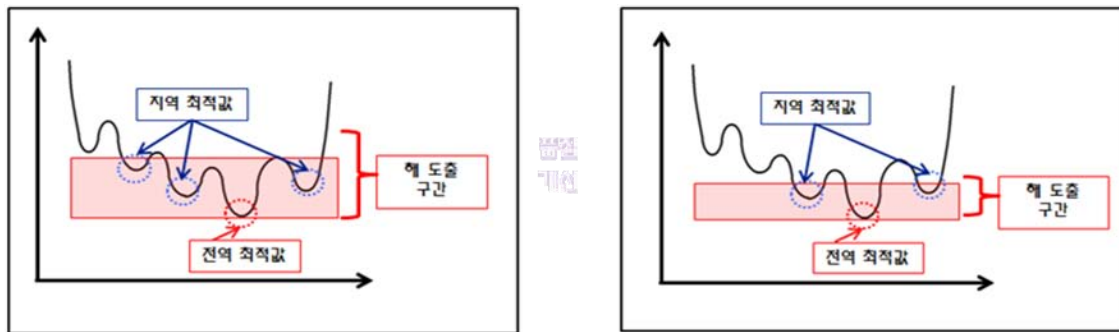


Fig 8 Improvement target of clustering algorithm

이러한 개선 목표에 따라 본 논문에서 제안하는 DSM 기반 Clustering 알고리즘의 특징을 정리해 보면 다음과 같다.

1. 병합식 알고리즘을 변경하여 사용한다. 일반적인 병합식 알고리즘은 정해진 Cluster 개수(N)가 될 때까지 Cluster 간의 상관 관계 점수를 평가하여 Cluster들을 N개로 병합해

나가는 알고리즘이다. 그러나 PBS를 구축하는 문제는 Cluster 내 객체간 상관 관계와 Cluster 간 상관 관계를 동시에 고려해야 한다. 따라서 본 알고리즘은 Thebeau의 목적함수(Cluster 내 객체간 상관 관계 점수와 Cluster 간 상관 관계 점수를 동시에 평가할 수 있음)를 사용하여 각 Clustering 단계마다 목적함수로 평가하여 다음 단계로 Clustering을 진행할지 말지를 결정하도록 한다. 목적함수는 아래 수식 2와 같다.

$$\text{Total cost} = \sum_{k=1}^N \text{In_point}_{\text{cluster}_k} \times \text{Cl_size}(\text{cluster}_k) + \sum \text{Out_point} \times \text{DSM}_{ij} \quad (2)$$

2. Bid 함수에 의해 그룹핑될 Cluster를 선택한다. Thebeau 알고리즘의 경우는 랜덤 함수에 의해 무작위로 객체를 선택한 후, 평가척도로 평가하여 새로운 Cluster를 생성할지 아니면 기존 Cluster의 멤버가 될지를 결정하였다. 그러나 이러한 무작위 추출법은 Clustering 대상 객체가 많아질수록 해 도출 구간이 넓어지며 작은 횟수의 반복 수행으로는 좋은 해를 기대하기 어렵다는 단점을 가진다. 따라서 모든 Cluster들 간의 상관 관계 점수를 평가하여 가장 높은 점수를 가지는 두 Cluster를 하나의 Cluster로 병합(agglomeration)하는 방법을 사용하였다. 이렇게 할 경우 큰 size의 Cluster로 모두 병합될 우려가 있기 때문에 상관 관계 점수를 Cluster size로 나눈 값을 척도로 사용한다. 알고리즘에 사용된 Bid 함수는 아래 수식 3과 같다.

$$\text{Bid}(G_k, G_t) = \frac{\sum_{i \in G_k \text{ and } j \in G_t} (\text{DSM}(i, j) + \text{DSM}(j, i))}{|G_k| + |G_t|}$$

- G_k : k 번째 클러스터의 객체 집합
- G_t : t 번째 클러스터의 객체 집합
- i, j : 객체 또는 DSM 요소

(3)

3. 각 Clustering 단계마다 목적함수로 평가하여 목적함수의 값이 다시 증가되는 시점에서 병합을 중지한다. 따라서 각 Clustering 단계마다 목적함수로 평가하여 목적함수 값이 감소하다가 다시 증가하는 시점에서 병합을 중지한다.

위 3 가지 특징을 가지는 새로이 제안되는 DSM 기반 Clustering 알고리즘의 전체적인 프로세스를 그림으로 표현하면 아래 그림 9 와 같다. 그림 10 은 DSM size 5(대상 객체 개수 5)인 경우를 예로 들어서 그림 9 의 프로세스를 따라 어떻게 Clustering 이 수행되는지를 설명한

그림이다.

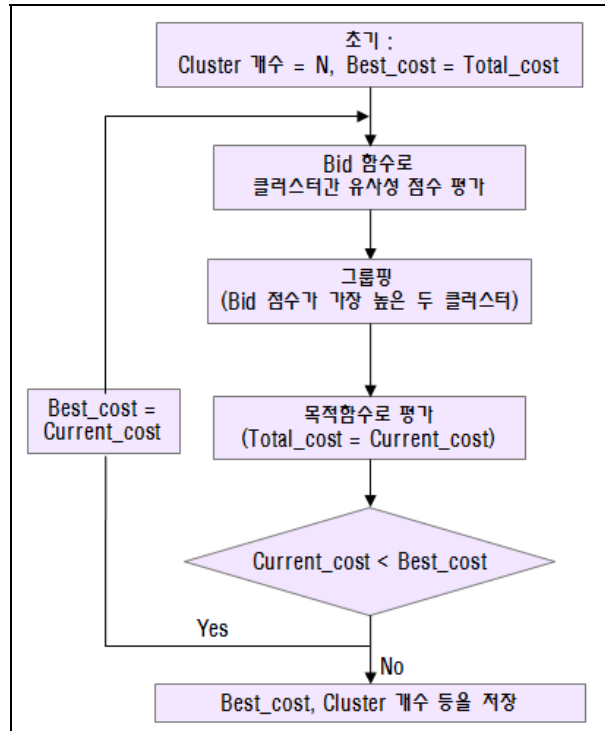


Fig 9 Flowchart of the improved clustering algorithm

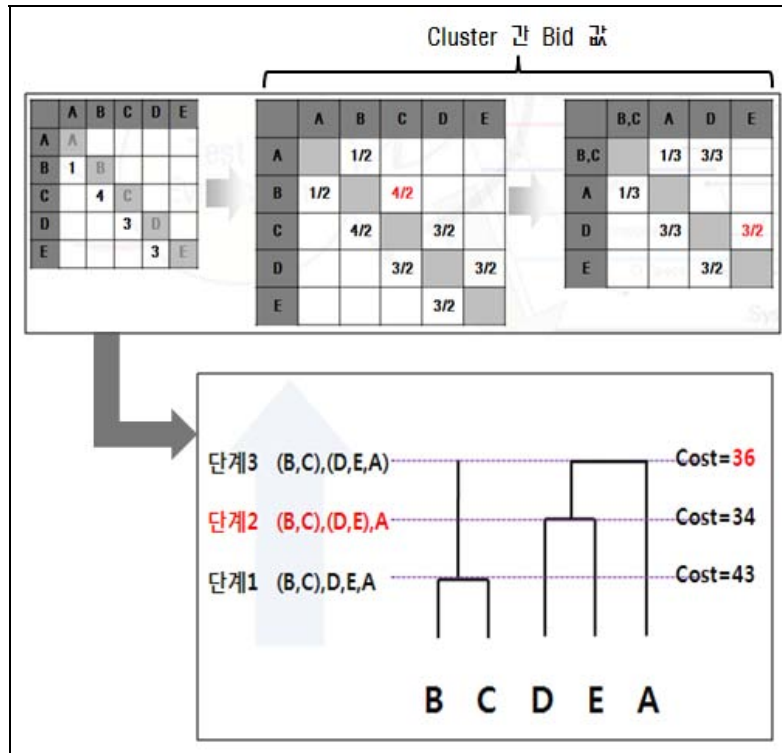


Fig 10 Example of clustering (dendrogram)

개선된 Clustering 알고리즘의 성능 테스트. 개선된 DSM 기반 Clustering 알고리즘의 성능을 시험하기 위해서 동일한 DSM 을 대상으로 Clustering 분석을 수행한 후, Clustering 결과를 비교 분석하였다.

실험 방법을 정리해 보면 아래와 같다.

- 대상 DSM

- DSM size: size 15(30 개), size 30(30 개), size 45(30 개)
- DSM 내 연관 관계 개수: size 15(25 개), size 30(50 개), size 45(75 개)

- 실험 횟수

DSM size 15, size 30, size 45 각각에 대해

- 기존 프로그램: 각 모델 당 10 회씩 수행
- 개선 프로그램: 각 모델 당 1 회 수행

- 결과 비교

- 기존 프로그램: 각 모델 당 10 회 반복 수행으로 얻은 결과값 중 최소값
- 개선 프로그램: 각 모델 당 1 회 수행으로 얻은 결과값

이런 실험을 통하여 아래 표 4 와 같은 결과를 얻었다. 아래 표 4 는 각 DSM size 별로 30 개의 모델에 대해 기존의 DSM 기반 Clustering 알고리즘으로 10 회 반복 수행하여 얻은 최소값과 개선된 Clustering 알고리즘으로 1 회 수행하여 얻은 결과값을 기록한 것이다.

동일한 DSM 을 대상으로 기존 알고리즘의 결과값을 X1 이라 하고 개선된 알고리즘의 결과값을 X2 라고 할 경우에 개선 알고리즘이 기존 알고리즘에 비해 얼마나 성능이 향상되었는지를 나타내는 척도를 “성능향상도(PI)”라 하고 아래 식 3-4 와 같이 정의한다.

$$PI = \left(1 - \frac{X_1}{X_2}\right) \times 100 \quad (4)$$

이런 정의에 따라 표본 30 개의 성능향상도(PI)들에 대한 표본 평균과 표본 표준편차는 수식 5 와 수식 6 과 같다. 이 경우 성능향상도 모평균의 95% 신뢰 구간은 수식 7 과 같다.

$$\bar{PI} = \frac{\sum_{i=1}^n PI_i}{n} \quad (5)$$

$$S_{PI} = \sqrt{\frac{\sum_{i=1}^n (PI_i - \bar{PI})^2}{n-1}} \quad (6)$$

$$\bar{PI} - t_{n-1,0.025} \frac{S_{PI}}{\sqrt{n}} \leq \mu_{PI} \leq \bar{PI} + t_{n-1,0.025} \frac{S_{PI}}{\sqrt{n}} \quad (7)$$

위 식 7 을 활용하여 DSM 각 size 별로 신뢰수준 95%에서 성능향상도의 모평균을 구해 보면 아래와 같다.

- DSM size 15×15 일 경우:

$$-3.832 \leq \mu_{PI} \leq 0.538$$

- DSM size 30×30 일 경우:

$$7.395 \leq \mu_{PI} \leq 11.993$$

- DSM size 45×45 일 경우:

13.516 ≤ μ_{PI} ≤ 17.155

위 결과에서 보듯이 DSM size 가 비교적 작은 15 의 경우에는 개선 알고리즘의 결과값은 기존 알고리즘의 결과값과 비슷하지만 DSM size 가 30, 45 로 증가할수록 개선 알고리즘의 결과값이 기존 알고리즘의 결과값보다 약 9.6%, 15.3% 적은 값(성능향상도 = 9.6, 15.3)을 도출하는 것을 알 수 있다.

따라서, 위 결과로부터 DSM size 가 커질수록 개선 알고리즘의 결과값이 기존 알고리즘의 결과값보다 더 좋은 품질의 근사 최적값을 구한다는 것을 알 수 있다.

결론

본 논문에서는 기존의 DSM 방법론의 문제점을 분석하고 이를 개선하는 방안을 제시하였다. 연구 결과를 요약해 보면 다음과 같다.

1. PBS를 구축하는데 사용되는 기존의 DSM Clustering 분석 방법론의 문제점을 분석하고 이를 개선하는 방안을 제시

PBS 를 구축하는데 사용되는 기존의 DSM Clustering 방법론은 외부 시스템을 분석 대상에 포함하지 않음으로써 올바른 PBS 결과물을 얻을 수 없었다. 또한 시스템 구성품들 간의 상관 관계는 오로지 동적인 기능을 수행하는 구성품들 간의 관계만을 표현할 수 있으며, 정적 기능을 수행하는 구성품들 간의 상관 관계는 표현할 수 없었다. 이러한 문제점을 해결하기 위해서 본 논문에서는 개발 대상 시스템의 정황을 분석한 후, 외부 시스템을 식별하여 이를 Clustering 분석 대상 요소로 포함시키고 또한, 정적 기능을 수행하는 구성품들 간의 상관 관계를 표현할 수 있는 새로운 상관 관계 카테고리 생성하였다.

2. PBS를 구축하는데 사용되는 기존의 DSM 기반 Clustering 알고리즘의 문제점을 분석하고 이를 개선하는 방안을 제시

Thebeau 알고리즘은 Clustering 개수가 작은 경우에는 랜덤함수를 이용하여 근사 최적값이 아닌 최적값을 구할 수 있다는 장점을 가지지만 Clustering 개수가 큰 경우에는 반복 계산 횟수가 증가하게 되므로, 제한된 반복 계산 횟수 내에서 최소값을 선택하게 할 수 밖에 없다. 따라서 Thebeau 알고리즘은 Clustering 개수가 큰 경우에는 근사 최적값을 도출하게 되며, 그 해의 품질이 좋지 못하다는 단점을 가진다. 이러한 문제를 해결하기 위해 본 논문에서는 병합식 Clustering 알고리즘과 Clustering 대상 객체를 선택하기 위한

평가함수(Bid 함수)를 사용함으로써 빠른 시간 내에 좋은 품질의 해를 찾을 수 있는 방법을 제시하였다.

이와 같이 본 연구 결과는 PBS 구축에 활용되는 기존의 DSM 기반 Clustering 방법론의 문제점을 개선함으로써 방법론의 성능이나 성숙도 측면에서 많은 발전을 이루었지만 Clustering 알고리즘의 성능 개선 부분에 관해서는 추가적인 연구가 필요할 것으로 사료된다. 본 논문에서 제시한 Clustering 알고리즘의 경우에 동일한 Cluster 간 상관 관계 점수를 가질 때 하나를 선택할 수 밖에 없으며 이러한 이유로 최적값이 아닌 근사 최적값을 도출하게 된다. 따라서 동일한 상관 관계 점수를 가질 때 이를 다시 세부적으로 분류할 수 있는 평가 척도를 개발한다면 보다 좋은 품질의 해가 도출될 것으로 사료된다.

참고 문헌

- [1] 박한샘, “진화연산과 적응적 α -cut 기반 평가를 이용한 유전자 발현 데이터의 퍼지 클러스터 분석”, *정보과학회, 소프트웨어 및 응용* 33 권 8 호, pp681, 2006
- [2] 임동순, “유사성 계수에 의한 객체 클러스터링 문제에서 최적 클러스터 수와 클러스터 멤버 결정”, *대한산업공학회 추계학술대회*, pp.488-495, 2002
- [3] 한용학, 패턴인식 개론, 한빛 미디어, 서울, pp 183, 2006
- [4] G.Fung, “A Comprehensive Overview of Basic Clustering Algorithms”, June 22, 2001
- [5] James N. Martin, *Systems Engineering Guidebook: A Processes for Developing Systems and Products*, CRC Press, USA, pp. 216, 1996
- [6] Manuel E. Sosa, Steven D. Eppinger, Craig M. Rowles, “Identifying modular and integrative systems and their impact on design team interactions”, *ASME*, volume 125, June 2003
- [7] McCord, Kent R. and Eppinger, Steven D., "Managing the Integration Problem in Concurrent Engineering", M.I.T. Sloan School of Management, Cambridge, MA, Working Paper no.3594, 1993.
- [8] Ronnie E. Thebeau, “Knowledge Management of System Interfaces and Interactions for Product Development Processes”, MIT, 2001

- [9] Soo-Haeng Cho and Steven D. Eppinger, “A Simulation-Based Process Model for Managing Complex Design Projects, *IEEE Transactions on Engineering Management*, Vol. 52, No. 3, 2005
- [10] Steven D. Eppinger, “Identifying Modular and Integrative Systems and their Impact on Design Team Interactions”, *ASME*, vol. 125, 2003
- [11] Steven D. Eppinger, Murthy V. Nukala, Daniel E. Whitney, “Generalized models of design iteration using signal flow graphs”, MIT, working paper number 3866, 1996
- [12] Tyson R. Browning, “Applying the design structure matrix to system decomposition and integration problems: a review and new directions”, *IEEE Transactions on Engineering Management*, Vol. 48, No. 48, pp. 299-304, 2001
- [13] Tyson R. Browning, “Integrative Mechanisms for Multi-team Integration: Findings from five case studies”, *systems engineering*, Vol.1, pp 95~112, 1999

Improvements to DSM Clustering Approach for PBS Construction

Chan Mook Kim

Department of Systems Engineering,
Graduate School of Ajou University,
Republic of Korea

kcmtg@nate.com

Young Won Park

ywpark@ajou.ac.kr

Abstract: A top-down scenario based approach is a typical method generating the product breakdown structure (PBS) in systems engineering. While it takes time and effort to perform the scenario based PBS construction, it is the safest approach not to omit any functions. On the other hand, a design structure matrix (DSM) approach to PBS construction is a bottom-up approach based on modularization concept that provides a merit of quick PBS construction while the sanity of its results may not be assured.

Accordingly, this study presents trade-off results between the two approaches to uncover problems in the DSM-based clustering method and proposes improvement approaches to enable the interfaces with external systems, static elements, as well as a new capability to handle the large number of clustering elements.

Introduction

The PBS is a result of systems design that illustrates system components by hierarchical structure. For building PBS, there is a scenario based PBS method that is proposed by system engineering. The scenario based method is the method that looks for the requirements function for system by describing usage of system according to time flow. This is a top-down approach analysing from system level to sub-component level. This method is particularly useful for developing new system. However, the scenario based method requires a lot of time and efforts. It needs to make a scenario for each level and compare/analyse various alternatives for implementing functions to physical components. In summary, the scenario based PBS method has an advantage that it would not omit any function since it is based on requirements, however, it has inconveniences that it requires a lot of time and efforts to develop.

On comparing, DSM Clustering based PBS method is more efficient and faster. It first selects components of appropriate level and then, identifies interface among selected

components. Based on modularization concept, it next performs grouping on those components that connected closely each other. This is a Bottom-up approach. Clustering is a grouping of highly related components, and performs fast by computer application with clustering algorithm. Therefore, this method can get a good result much faster than scenario based method. However, it has several problems since it is not verified over the time like scenario based method. Hence, this paper attempts to compare scenario based method and DSM clustering based method by applying to same railroad car system. The problems of DSM clustering based method will be addressed and the solutions for these problems will be discussed. Furthermore, since the result of DSM clustering based method depends a lot on clustering algorithm, the study for improving this method should accompany with clustering algorithm improvement. Therefore, this study aims to propose more efficient DSM based clustering algorithm through analyzing existing clustering algorithms.

Problems & improvements of DSM based Clustering method

This section shows the result of scenario based system architecting method and DSM based clustering method applying to same target system; railroad car system. PBS results from each method will be compared in order to find out the problem of DSM based clustering method, and we will discuss how to improve.

PBS construction by Scenario based system architecting method. The system design by systems engineering starts from requirements analysis. Martin defined the design process like <Fig 1>. The system design first collects requirements from stakeholders, and makes scenarios based on this. The reason of making scenario is to include every movements of system and users according to time flow so as to induce the system function without any omission. The function here based on collected requirements, and scenario contains both dynamic and static functions.

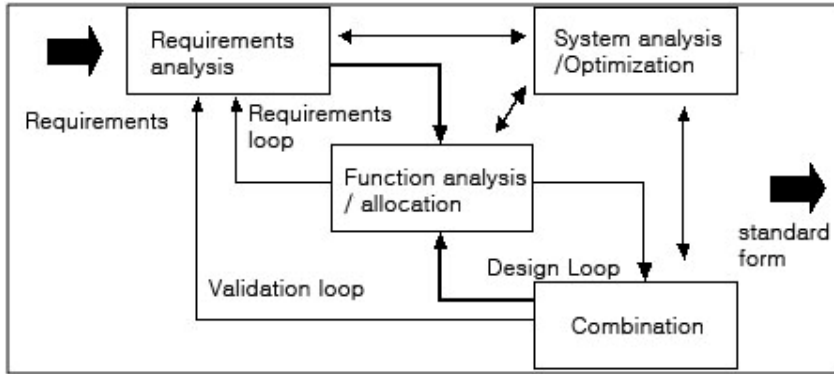


Fig 1 System design process

Martin defined 'Function' as a task or activity in order to obtain the desirable outcome, or capability of system and system elements. This function can be static or dynamic. In here, dynamic function is a function providing output through processing/handling/converting a certain input, while static function is a function without processing/handling/converting. These static and dynamic functions perform a system analysis and optimization, and the sub system is determined at compounding stage.

Scenario based PBS construction method is a "Top-down" approach as it is decomposed from system level to sub system level by repetitive loop of requirements analysis/function analysis/composition activities. PBS construction for railroad car system has to define first the system boundary and then system and external system. Therefore, the system context diagram of railroad car system can be drawn like <Fig 2>.

As have seen in <Fig 2>, the external system of railroad car system includes driver, passenger, power system and signal/communication system. After defining system and external system, system scenario can be made. FFBD (Function Flow Block Diagram) of system scenario about passenger and railroad car system is presented in <Fig 3>.

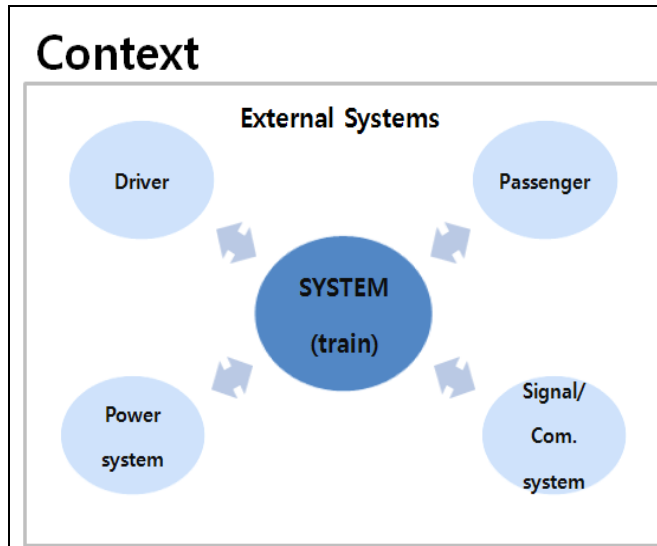


Fig 2. The context diagram of railroad car system

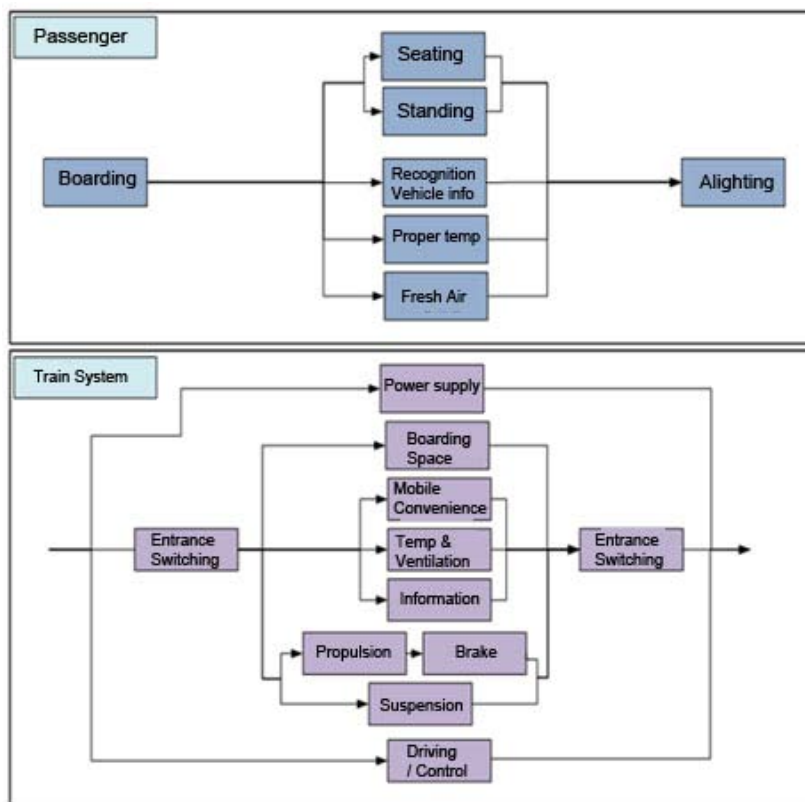


Fig 3 The system scenario of railroad car system

The system function can be defined based on scenario such like <Fig 3>. After that, it

goes through composition stage and defines sub system. As it is repeated process in order to generate PBS, we can build PBS such like shown in <Table 1>. As you can see from Table1, the design process that is on system level defined 5 of sub systems (level 1) and each sub system has sub components(level 2).

Level 0	Level 1	Level 2
Railroad car system	Driving & Control	Total Control Device
		Auto car control device
		Auto car driving device
		Driving cabin facility
	Passenger Cabin	Chassis
		Window
		Entrance
		Interior design
		Aisle connection between cars
	Power Supply	Pantograph
		Sub power supply
	Facility	Chair
		Passenger cabin handle
		Passenger guide facility
		Air-conditioner/heater & ventilation device
	Travelling	Bogie Frame
		Air compressor
		Connector
		Hauling device
		Suspension equipment
		Propulsion equipment

		Brake equipment
		Power train

Table 1 The PBS of railroad car system (scenario based)

PBS construction by DSM based Clustering. Unlike to scenario based architecting method, DSM based clustering method is performed by “Bottom-up” approach. The first thing has to be done is “select component of appropriate level”. For railroad car system, total 30 components were chosen (see Table 2).

No.	Components	No.	Components
1	Total control device	16	Chair
2	Auto car control device	17	Cabin handle
3	Auto car driving device	18	Information guide device
4	Driving Cabin facility	19	HVAC
5	Under Frame	20	Bogie Frame
6	Side Frame	21	Hauling device
7	Roof Frame	22	Air compressor
8	Bottom Platform	23	Connector
9	Window	24	Suspension device
10	Entrance	25	Inverter
11	Interior design	26	Propulsion Device
12	Aisle connection between cars	27	Brake
13	Pantograph	28	Wheel
14	Charger	29	Power transmission device
15	Battery	30	Wheel-axle

Table 2 Target components for DSM clustering

When the components are selected, these are registered in DSM (component-based DSM). On the Component-based DSM, correlations among components can be divided into 4; energy exchange, information exchange, material exchange and physical interface.

With respect to 4 types of correlation, we give the score 0(no correlation), 5(a little of correlation) and 10(strong correlation) according to strength of correlation among components. As correlation scores are given like <Fig 4>, clustering analysis is now performed. The <Fig 5> shows the result of upper components cluster in terms of high correlation, after analysing clustering analysis of <Fig 4>.

Below <Table 3> gives PBS of railroad car system based on cluster result of DSM clustering analysis. The case of cluster 2 cannot be found its representative upper function for all the function of internal components.

	통관제	자동 제	자동 제	연터프	사이드	지붕 프	문단제	창문	출입문	내장	지랑간	입전기	충전기	배터리	의자	객실	정보	HVAC	대자 프	건인장	공기압	연결장	현가장	인버터	추진모	브레이크	휠	동력전	운축	
종합제어장치	10	10	5					5			5	5	5					5	5			5	5	10	5	5				
자동열차제어장치	10		10	5																				5						
자동열차운전장치	10	10	5					5																5						
운전실 열비	10	10	5																					5						
언더프레임					10	10																								
사이드 프레임				10	10	10	5	5	5																					
지붕 프레임				10	10	10												5												
끝단제				10	10	10					5																			
창문					5																									
출입문		5	10	5																			10		5					
내장					5														5											
차랑간 통로연결막							5																							
집전기				5																								5		
충전기		5																							10					
배터리														20																
의자										5																				
객실 손잡이										5																				
정보 안내 장치			5	5						5																	5			
HVAC		5		5		5				5																				
대자 프레임																														
건인장치																				10		10	10	10		10	10	10		
공기압축기		5																		10										
연결장치																				10		10								
현가장치																				10		10								
인버터		5	10	10								10																20		
추진모터																				10							20			
브레이크		5	10																	10		10								
휠																												10	10	10
동력전달장치																				10						10		10	10	
운축																											20	20		

Fig 4 The DSM of railroad car system

IV. 클러스터링(clustering) 결과																																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30		
언더프레임	10	10	10																													
사이드 프레임	10	10	10	5	5										5																	
지붕 프레임	10	10	10							5																						
끝단체	10	10	10				5																									
창문		5																														
내장		5								5																						
차량간 통로연결막				5																												
의자						5																										
객실 손잡이						5																										
HVAC			5			5					5				5				5													
종합제어장치										5	10	10	5	5	5	5	10	5			5		5	5	5	5	5					
자동열차제어장치										10	10	10	5					5														
자동열차운전장치										10	10	5	5	5				5														
운전실 설비										10	10	5						5														
출입문		5									5	10	5					5					10									
집전기													5					5														
정보 안내 장치					5								5	5					5													
인버터										5	10	10			10				20													
추진모터																			20		10											
대차 프레임																			10	10	10	10	10	10	10					10		
견인장치																				10	10	10	10	10	10							
공기압축기										5										10	10	10	10	10	10							
연결장치																				10	10	10	10	10	10							
현가장치																				10	10	10	10	10	10							
브레이크										5	10									10	10	10	10	10	10							
충전기											5									10												
배터리																															20	
휠																															10	10
동력전달장치																															10	10
윤축																															20	20

Fig 5 The result of clustering for railroad car components

Level 0	Level 1	Level 2	Level 3
Railroad Car system	Cluster 1 (Cabin)	Chassis	Under frame
			Side frame
			Roof frame
			Partition frame
		Window	
		Entrance	
		Interior	
		Aisle connection between cars	
		Chair	
	Handle		
	HVAC		
	Cluster 2	Driving & control	Total control device
			Auto car control
			Auto car driving device
			Driving facility
		Entrance	
		Pantograph	
Information guide device			
Propulsion		Inverter	

			Propulsion motor
	Cluster 3 (Power Supply)	Charger	
		Battery	
	Cluster 4 (Power transmission)	Wheel	
		Power transmission device	
		Wheel-axle	
	Cluster 5 (Travelling)	Bogie Frame	
		Hauling equipment	
		Air compressor	
		Connector	
		Suspension equipment	
	Brake		

Table 3 The railroad car's PBS resulted from DSM clustering

Problems and Improvements of DSM based Clustering Method. We can find the biggest difference between scenario based system architecting method and DSM clustering analysis at the category of “additional facility”.

First of all, when we look at PBS resulted from DSM clustering analysis, “information guide device” and “HVAC” cannot form any cluster as “additional facility”, hence it belongs to Cluster 1(Cabin). The main reason of this difference is that existing DSM based clustering method restricts DSM elements (analysis target) only to target system. The “information guide device” offers information about current station and next station to passengers, while “HVAC” provides fresh and proper temperature air to passengers. However, since existing DSM method restricts DSM elements to clustering analysis target system, it cannot mark information exchange between passenger and information guide device, and material exchange between passenger and HVAC. Therefore, there is no chance to bind passenger, HVAC and information guide device to one cluster.

Secondly, “Chair” and “Cabin handle” also cannot form a cluster, but belong to Cluster 1(Cabin). The reason of this difference is because existing DSM based clustering method does not consider the correlation with external system, and with existing 4- correlation (energy exchange, information exchange, material exchange and physical interface) it cannot describe the correlation among components those perform static

functions. The function of both “Chair” and “Cabin handle” is to provide convenience and safety to passengers. These two items do not result output through processing with any input. Therefore, according to existing relation category, since “Chair” and “Cabin handle” does not have any energy, information and material exchange with passengers and no relation indication, Passenger, Chair and Cabin handle cannot be bound to one cluster. However, although Chair and Cabin handle does not result output through processing with any input, it certainly provides convenience to passengers. We need a new correlation category in order to form a cluster of static function components (i.e. chair and cabin handle) and passengers.

Finally, we propose two methods to solve the problems of existing DSM based Clustering method.

1. When we construct PBS using by DSM based Clustering method, performing ‘clustering’ including external system as target elements.

Some of clustering object components has stronger correlation with external system. Hence, it can be performed ‘clustering analysis’ after including external system to clustering object components. The cluster that is a result of clustering analysis may contain external system. In this case, except external systems, the rest of components can be regarded as actual cluster components. For example, it integrated the components of railroad car system and external system and then analysed. The result shows that Cluster 1 is composed of passenger, information guide device and HVAC, however except passenger, only information guide device and HVAC is regarded as Cluster 1’s components.

2. Generating new correlation category “service” that is able to describe correlation among components those perform a static function.

Though static function components do not process input and provide output, these have a certain purpose, elements and correlation for accomplishing the goal. Therefore, we generate new correlation category “service” in order to describe the correlation among static function components. These static function components

mainly work for passenger security or convenience, thus it is reasonable to categorise to “service”.

Problem and Solution of DSM based clustering algorithm

The core of DSM method is clustering analysis. The clustering analysis chooses the best output among the outputs came from numerous iterative works. Therefore, clustering analysis is difficult to calculate by manual; all clustering analysis is performed by computer auto program. In summary, the core of DSM method is clustering analysis, clustering analysis is performed by auto program, and therefore, clustering analysis relies on clustering algorithm.

In this chapter, we analyse the existing clustering analysis program in order to find the problem and propose our algorithm that can solve this problem.

Characteristic of clustering algorithm. The characteristics of clustering problems are shown below.

1. Receiving only correlation between two objects. It cannot describe characteristic vector for each object. It is only provided correlation between two objects.
2. One object belonging to only one cluster. In the PBS, one component belongs to one upper component (or subsystem) and cannot belong to various upper components.
3. Considering correlation of intra-cluster and inter-cluster simultaneously. It is same as Lim's $\lambda=0.5$ criteria method.
4. Not being able to fix the number of cluster in advance. The clustering algorithm of parameter approach usually starts with fixed number of cluster at the initial stage. There are few algorithms those start running without knowing number of cluster. However, these algorithms are difficult to apply in PBS construction. Therefore, we discuss this issue in the next chapter.
5. Searching approximate optimal value rather than optimal value. The possible solution case in the clustering problem can be expressed like formulation (1). As have seen in below formulation, when we do not know the number of cluster, the number of alternative increases. Hence, most of researches consider approximate optimal value within limited time rather than finding optimal value.

$$\sum_{M=1}^N \left(\frac{1}{M!} \sum_{j=1}^M (-1)^{M-j} {}_M C_j \times j^N \right)$$

- M: Cluster Coefficient
- N: No. of Cluster
- ${}_M C_j$: Combination

The case of N known

The case of N unknown

(1)

The types of existing clustering algorithm. The problem like PBS construction which only knows correlation between object and do not know the number of cluster at initial stage, generally applies partial optimization by iterative experiments. As it increases cluster number, partial optimization is performed at each stage. The clustering algorithm process by iterative experiments is shown in <Fig 6>. As seen in <Fig 6>, when we increase the cluster number, the objective function value gradually increases (or decreases in terms of objective function) and decreases (or increases) when it reaches at a certain level. At this point, we assume it is an optimal cluster number. Therefore, this clustering algorithm by a repetitive test is time costly.

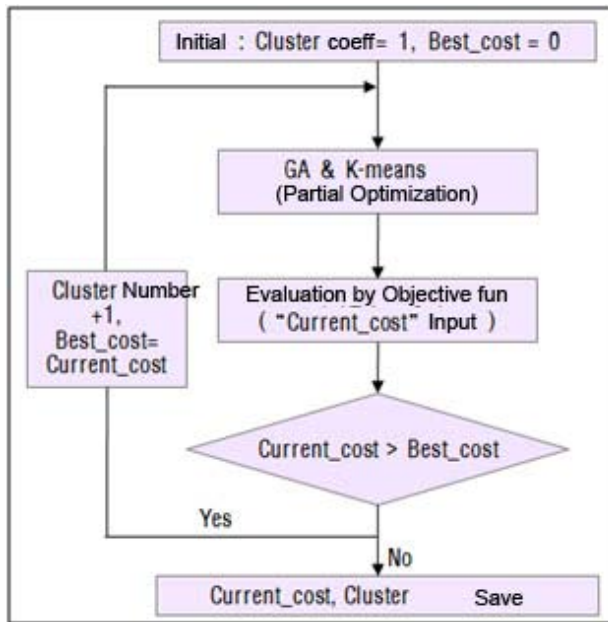


Fig 6 Clustering algorithm by using a repetitive test

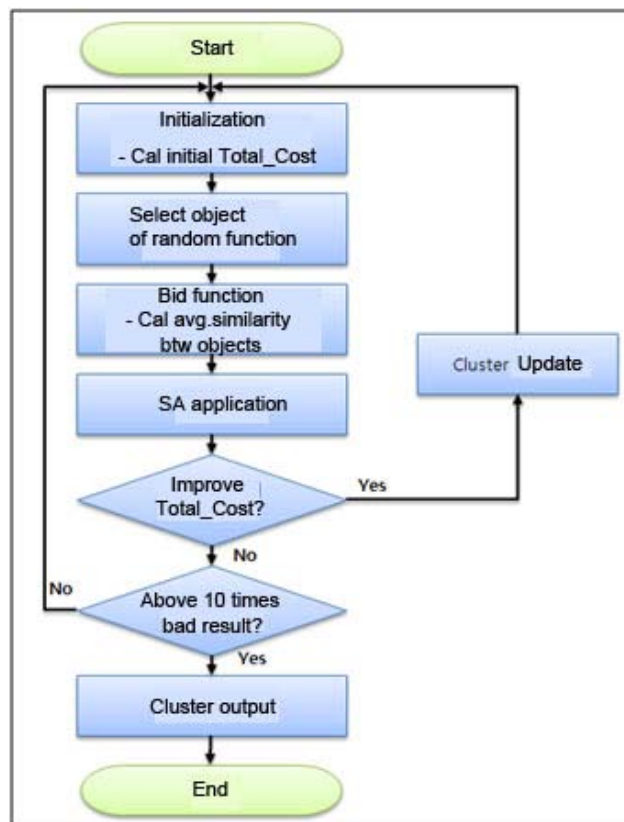


Fig 7 Flowchart of the Thebeau's clustering algorithm

Ronnie E. Thebeau has developed DSM based clustering algorithm by improving Fernandez’s algorithm. The characteristics of Thebeau algorithm is that it uses random function for selecting clustering objects. If we set any rule for selecting clustering object, it may not be able to find optimal value. Thus, by using random object abstract method, we can benefit to find optimal value rather than approximate optimal value. Meanwhile, as clustering objects increases, the repetitive calculation time for obtain optimal value gets longer. The thebeau algorithm process is given in <Fig 7>.

Problem & Solution of existing clustering algorithm. As we discussed, the repetitive test of clustering algorithm has a time costly disadvantage. Thebeau algorithm also has same problem that when the clustering object is large, it needs to perform numerous iterative calculation in order to obtain an optimal solution, although it has a merit of random object abstract. It means Thebeau algorithm is also not able to know when the optimal value can be appeared with large clustering object; therefore it has no choice but to choose minimum value among repetitive test results.

Based on examination of algorithm, our study aims to develop the algorithm that enables to find an optimal value very fast, even if clustering object is large. The efficiency difference between Thebeau algorithm and proposed algorithm is shown in <Fig 8>.

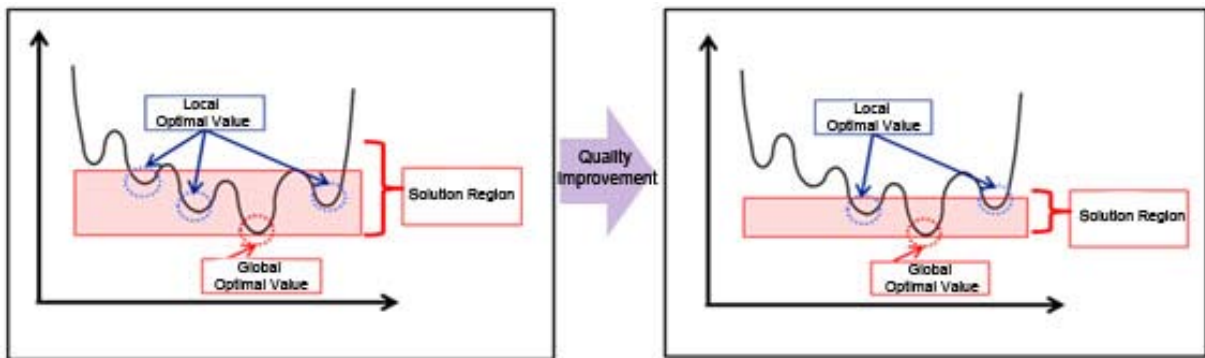


Fig 8 Improvement target of clustering algorithm

Based on our algorithm improvement goal, the characteristics of our proposed DSM based clustering algorithm can be summarized as shown in below.

1. **Adopting agglomerative algorithm.** The general agglomerative algorithm evaluates correlation between clusters until it reaches N cluster number; clusters are consolidated to N. However, for PBS construction problem, it has to consider the correlation of both inter-cluster and intra-cluster at the same time. Hence, our algorithm uses Thebeau's objective function (possible to evaluate correlation score of inter-cluster and intra-cluster simultaneously) as criteria for allowing procedure for next stage clustering. The objective function is seen in formulation (2).

$$\text{Total cost} = \sum_{k=1}^N \text{In_point}_{\text{cluster}_k} \times \text{Cl_size}(\text{cluster}_k) + \sum \text{Out_point} \times \text{DSM_size} \quad (2)$$

2. **Selecting cluster for grouping by Bid function.** The case of Thebeau algorithm chooses objects by random function and evaluates with criteria to decide whether generating a new cluster or adding to existing cluster. However, this random sampling method cannot provide a good solution when there is large number of objects. Therefore, we use agglomerative method; investigating correlation score among all clusters to combine the highest two clusters into one. In this case, all clusters could be merged into large size cluster. To protect this case, we adopt the criteria value that comes from dividing correlation score by cluster size. The Bid function we used in this study is presented in formulation (3).

$$\text{Bid}(G_k, G_t) = \frac{\sum_{i \in G_k \text{ and } j \in G_t} (\text{DSM}(i, j) + \text{DSM}(j, i))}{|G_k| + |G_t|}$$

- G_k : kth cluster's objects aggregation
- G_t : tth cluster's objects aggregation
- i, j : Object or DSM element

(3)

3. Each clustering stage is evaluated by objective function. When it reaches at the point that the value of objective function increases again, stop agglomeration.

The process of this newly proposed DSM based clustering algorithm can be described like <Fig 9>. <Fig 10> explains how clustering can be performed like process shown in <Fig 9> using the example of DSM size 5(target objects-5).

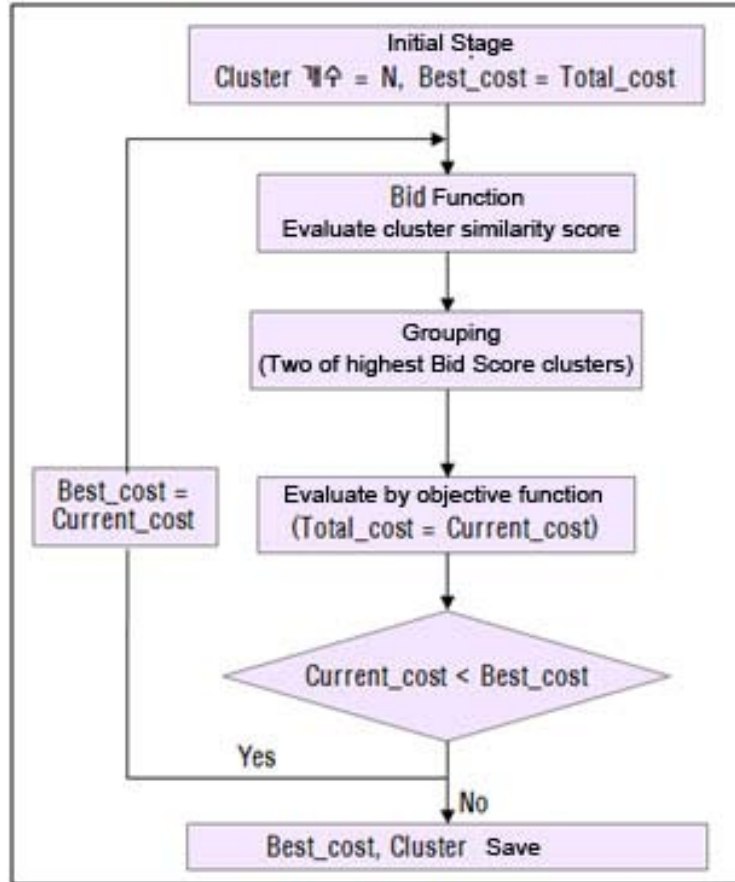


Fig 9 Flowchart of the improved clustering algorithm

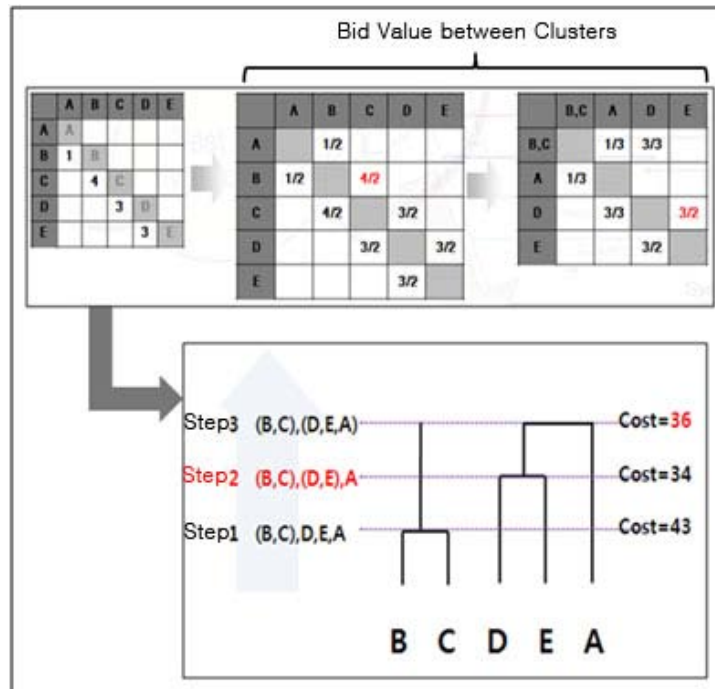


Fig 10 Example of clustering (dendrogram)

Efficiency Test of improved Clustering algorithm. For efficiency test of improved DSM based clustering algorithm, same DSM was tested with proposed algorithm and its clustering result was compared with existing one.

The test method is followed below.

- Target DSM
 - DSM size: size 15(30EA), size 30(30EA), size 45(30EA)
 - No.of correlation in DSM: size 15(25EA), size 30(50EA), size 45(75EA)
- Number of experiments

For DSM size 15, size 30 and size 45, respectively

 - Existing Program: Performing 10 times for each model
 - Improved Program: Performing 1 time for each model
- Result comparison
 - Existing Program: Minimum value among the results
 - Improved Program: Result from experiments

The results from this experiment are shown in <Table 4>.

Considering same DSM, let X_1 be a result value of existing algorithm, X_2 be a result of improved algorithm and PI(efficiency) be a index of improved efficiency. PI is defined with below formulation 4.

$$PI = \left(1 - \frac{X_2}{X_1}\right) \times 100 \quad (4)$$

According to this definition, the standard mean and deviation of 30-PI sample can be calculated by formulation 5 and 6. The 95% confidence level of PI's population mean can be computed by formulation 7.

$$\overline{PI} = \frac{\sum PI_i}{n} \quad (5)$$

$$S_{PI} = \sqrt{\frac{\sum (PI_i - \overline{PI})^2}{n-1}} \quad (6)$$

$$\overline{PI} - t_{n-1,0.025} \frac{S_{PI}}{\sqrt{n}} \leq \mu_{PI} \leq \overline{PI} + t_{n-1,0.025} \frac{S_{PI}}{\sqrt{n}} \quad (7)$$

The population mean of efficiency improvement at the level of confidence 95% for each size of DSM can be calculated as followed.

- Case of DSM size 15×15:

$$-3.832 \leq \mu_{PI} \leq 0.538$$

- Case of DSM size 30×30:

$$7.395 \leq \mu_{PI} \leq 11.993$$

- Case of DSM size 45×45:

$$13.516 \leq \mu_{PI} \leq 17.155$$

As have seen in above result, the case of DSM size 15 which is relatively small shows similar result value as existing algorithm result, however as DSM size increases to 30 and 45, improved algorithm performs better; $H_{PF} = 9.6$ and 15.3, respectively.

Therefore, we found that as DSM size gets larger, output of improved algorithm can find better approximate optimal value than existing algorithm.

Conclusion

This study analyzed the problem of existing DSM method and proposed the solution. The summary of this study is as followed.

1. Investigated the problem of existing DSM Clustering analysis method and proposed its solution.

Existing DSM Clustering method usually does not consider external system as an analysis target. Therefore, there is no guarantee that the PBS result from DSM clustering is correct. Furthermore, existing method is limited to describing correlation among components those perform the dynamic functions; not able to describe correlation among components those perform the static functions. In order to solve these problems, this study included external system into clustering analysis target and generated new correlation category that is able to explain the correlation among components those execute static functions.

2. Investigated the problem of existing DSM Clustering algorithm and proposed its solution.

Thebeau algorithm has an advantage that when the number of clustering is small, it is possible to obtain the optimal value (not approximate optimal) using random function. However, when it has a large number of clustering, it has no choice but to choose minimum value within limited iterative calculation as its iterative calculation number increases. Thus, Thebeau algorithm draws approximate optimal value for the case of large number of clustering, which means the quality of solution is not good. To solve this problem, we propose 'agglomerative

clustering algorithm' and bid function in order to choose clustering object so as to find a good solution fast.

Although this study solved the problem of existing DSM based Clustering method and contributed to efficiency and maturity of method, it still needs to study on improvement of clustering algorithm. The proposed clustering algorithm in this study only can choose one when there is tie correlation scores. Hence, it results approximate optimal value not optimal value. Therefore, if we can develop the criteria to classify the tie score case, we will be able to achieve the good quality solution.

References

- [1] Han Saem Park, "Fuzzy cluster analysis of Gene Expression Profiles Using Evolutionary Computation and Adaptive α -cut based Evaluation", Korea Institute of Information Scientist and Engineers, Software & application 33, Vol.8, pp681, 2006
- [2] Dong Soon Lim, "Optimal number of cluster and cluster member decision for object clustering problem by similarity coefficient ", Korea Institute of Industrial Engineers, *autumn conference*, pp.488-495, 200
- [3] Yong Hak Han, Theory of Pattern Recognition, Hanbit Media, Seoul, pp183, 2006
- [4] G.Fung, "A Comprehensive Overview of Basic Clustering Algorithms", June 22, 2001
- [5] James N. Martin, Systems Engineering Guidebook: A Processes for Developing Systems and Products, CRC Press, USA, pp. 216, 1996
- [6] Manuel E. Sosa, Steven D. Eppinger, Craig M. Rowles, "Identifying modular and integrative systems and their impact on design team interactions", *ASME*, volume 125, June 2003

- [7] McCord, Kent R. and Eppinger, Steven D., "Managing the Integration Problem in Concurrent Engineering", M.I.T. Sloan School of Management, Cambridge, MA, Working Paper no.3594, 1993.
- [8] Ronnie E. Thebeau, "Knowledge Management of System Interfaces and Interactions for Product Development Processes", MIT, 2001
- [9] Soo-Haeng Cho and Steven D. Eppinger, "A Simulation-Based Process Model for Managing Complex Design Projects, *IEEE Transactions on Engineering Management*, Vol. 52, No. 3, 2005
- [10] Steven D. Eppinger, "Identifying Modular and Integrative Systems and their Impact on Design Team Interactions", *ASME*, vol. 125, 2003
- [11] Steven D. Eppinger, Murthy V. Nukala, Daniel E. Whitney, "Generalized models of design iteration using signal flow graphs", MIT, working paper number 3866, 1996
- [12] Tyson R. Browning, "Applying the design structure matrix to system decomposition and integration problems: a review and new directions", *IEEE Transactions on Engineering Management*, Vol. 48, No. 48, pp. 299-304, 2001
- [13] Tyson R. Browning, "Integrative Mechanisms for Multi-team Integration: Findings from five case studies", *systems engineering*, Vol.1, pp 95~112, 1999