

## 정보시스템 전략적 계획을 위한 컴퓨터지원 통계적 접근방법

김진수<sup>1)</sup>, 황철언<sup>2)</sup>

### A Computer-Aided Statistical Approach to Strategic Information Systems Planning

*Strategic information systems planning (SISP) remains a critical issue of many organizations and also the top IS concern of chief executives. Therefore, researchers have investigated SISP practices and tried to improve a methodology. Among the various issues of SISP, systematically determining subject database groupings and fully automating the processes are important aspects. This study presents an alternate methodology using a statistical technique, a variable clustering approach, and systematic rules for determining database groupings, which can be fully automated. This methodology provides a strong theoretical justification as well as systematic and simple criteria for database groupings, enhanced interpretability of the output, and would be easy to include in CASE software application.*

---

1) 부산대학교 경영학과 조교수  
2) 부산대학교 경영학과 석사

## I. 서 론

정보시스템 전략적 계획(Strategic Information Systems Planning: SISP)은 기업의 장 단기 정보요구를 충족시킬 수 있는 정보시스템 계획을 수립할 수 있도록 도와주는 구조적 방법론으로서 장기적으로 효과적인 시스템 구축을 통한 투자효익의 극대화를 제공하는 장점을 가져다 준다. SISP의 이같은 중요성에 따라 정보시스템 분야 경영층뿐만 아니라 최고경영층에게도 지속적인 관심의 대상이 되고 있다[Dickson, Leitheiser, Wetherbe, and Nechis, 1980; Brancheau and Wetherbe, 1987; Niederman et. al., 1991; Earl, 1993].

SISP에 관한 지속적인 연구를 통하여 SISP를 위한 다양한 방법론이 제시되고 활용되고 있으나 대부분의 방법론이 기간이 장 기간 소요되고, 체계적인 면이 부족하며, 최고경영층의 참여가 부족하는 등 많은 문제점이 지적됨에 따라 방법론의 개선에 대한 연구가 시급한 실정이다. 하지만, 기존 연구는 주로 SISP를 위한 개념적 framework의 제시[W. King, 1978; 한재민, 문태수, 1993], 기존 방법론간의 장 단점 비교분석[Zachman, 1983; Pyburn, 1983; Sullivan, 1985; Hackathorn, 1988], 설문지를 통한 SISP 계획 및 실행상의 문제점 파악[Lederer and Sethi, 1988; Goodhue et. al., 1992] 등에 한정되고 있으며 정작 기존 방법론의 문제점을 해결할 수 있는 새로운

방법론의 개발에 관한 연구는 매우 미진한 실정이다.

SISP의 주요방법론으로는 IBM의 Business Systems Planning(BSP), James Martin의 Information Engineering(IE), Strategic Systems Planning, Method/1, Information Quality Analysis [Vacca, 1984], Critical Success Factors(CSF), Value Chain Analysis 등 다양한 방법론이 있다. 이 방법론 중에서 IBM에서 제시한 BSP는 경영지향적(business-oriented)인 접근방법으로서 하향식(top-down)계획과 상향식(bottom-up)구현 방법의 사용, 사용의 용이성, 체계적인 문서화를 지원함으로써 SISP 방법론 중에서도 가장 대표적인 방법론으로 널리 활용되고 있다[Lederer and Sethi, 1988].

하지만, BSP는 여러 장점에도 불구하고 정보시스템 체계수립과정의 임의적 처리(heuristic process)에 따라 최종산출물의 결과가 프로젝트 팀의 기술 및 창조적 능력에 좌우됨으로써 일관된 정보시스템 계획수립이 어려우며, 수작업 처리에 따른 과도한 문서화 부담, Computer-Assisted Software Engineering (CASE) tool과 같은 컴퓨터 지원 부족에 따른 계획기간의 장기화 등의 문제점이 지적되고 있다[Zachman, 1982; Crocket, Slinkman, and Eakin, 1989]. 이에 따라 Crocket et. al. [1989]은 정준상관 분석을 활용한 통계적 기법을 사용하여 CASE tool 환경에 적합한 새로운 체계적인

방법론을 제시하였지만 최종적으로 도출된 서브시스템의 정의를 위한 해석이 매우 어려운 단점을 나타냄에 따라 새로운 방법론의 개발이 절실한 실정이다.

따라서, 본 연구는 기존의 SISP 방법론 중에서 가장 널리 활용되는 BSP 방법론상의 문제점을 개선하여 효과적 SISP를 수행할 수 있는 새로운 방법론을 제시하고자 한다. 본 연구에서 제시한 방법론은 체계적이면서도 객관적인 절차 개발을 통한 일관된 결과 제시, CASE tool지원 가능한 방법론 제시를 통한 SISP과정의 자동화 지원과 문서화 부담 경감, 해석의 용이성에 따른 정보시스템 체계의 정확성 향상 등을 제공하는 장점을 제공한다.

방법론의 타당성 분석을 위하여 기존 BSP 방법론과의 비교연구 결과, 처리절차의 감소, 통계적 기법 사용을 통한 체계적 방법에 따른 자동화 가능, BSP 초기단계에서 파악된 프로세스 유지에 따른 해석상의 용이등이 제공됨으로써 새로운 방법론이 이론적으로나 실무적으로 매우 유용함이 입증되었다.

## II. 정보시스템 전략적계획에 관한 기존연구

SISP는 정보시스템 관리자가 직면한 가장 중요한 문제로서 그 목적은 정보시스템 구축을 위한 가장 적절한 대상(functional area)을 선정하고, 개발시스템의 설치에 대한 일정을 수립하는 것이라 할 수 있다 [Brancheau and Wetherbe,1987]. 이러한

SISP의 개념은 근래에 생성된 것으로 1970년대 후반 McLean과 Soden[1977]은 “시스템계획의 주요목적은 사용자와의 의사소통을 개선하고, 최고경영자의 지원을 증가시키고, 필요한 자원을 보다 효율적으로 예측하고 할당하며, MIS부문 개선을 위한 보다 많은 기회를 결정하고, 수익이 가장 높은 새로운 컴퓨터 응용프로그램(Application)을 확인하기 위한 것이다.” 라고 SISP의 개념을 정의하고 있다.

Lederer와 Sethi[1988]는 개념을 조금 더 확장하여 조직의 경영계획을 수행하여 결국에는 경영목표를 실현하기 위한 컴퓨터 응용프로그램의 포트폴리오를 확인하는 절차라고 하며, 여기에는 이들 응용프로그램을 지원하기 위한 데이터베이스와 시스템의 정의도 포함시키고 있다. 즉, 정보시스템 전략적계획은 주요 서브시스템과 서브시스템간의 관계 정립을 통하여 기업전체적인 관점에서 효과적인 정보시스템 구축을 위한 정보체계 (information architecture)를 제공함으로써 장기적으로 일관성있게 효과적인 시스템구축을 가능하게 한다.

기존의 SISP에 관한 연구를 살펴보면 크게, framework의 제시, SISP 방법론 간의 장단점에 관한 비교분석, 설문을 통한 SISP 계획 및 실행상의 문제점을 분석하는 것으로 파악할 수 있으며 구체적인 내용은 다음과 같다.

○Framework 제시: King[1978, 1988]은 조직의 전략집합을 MIS계획의 요소로 변

환할 것을 주장하고, 이를 위한 적용 가능한 일반적 절차로서의 접근법을 제시하였으며, 적절한 수행을 확인하기 위한 평가모델을 제시하였다. 또한 한재민 과 문태수 [1993]는 SISF를 위한 상황모델을 제시하면서 하나의 방법론만을 사용하기보다는 상황에 따라 BSP와 CSF 방법을 결합하여 사용하는 것이 보다 효과적인 SISF 방법론이라고 주장하였다.

○비교분석 : Zachman [1983]은 BSP와 Business Information Control Study (BICS)를 5 가지의 유사점과 차이점으로 열거 지적하고, 각 방법론이 가지는 장점과 단점을 제시하고 있다. Pyburn [1983]은 SISF의 접근법을 Personal-*Informal*, Personal-*Formal*, Impersonal-*Formal*의 세가지 틀로 구분하고, 설문문을 통해 응답 조직들의 사용 접근법을 상황적 특성으로 비교 연구하였다. Sullivan [1985]은 조직내 정보기술의 활용수준(운영통제-관리-전략)-*Infusion*-과 정보기술의 조직내 분산정도-*Diffusion*-에 따라 여러 SISF 방법론들이 사용되는 조직의 특성을 분류하고 있다. 즉, 방법론이 각 조직의 상황에 맞게 이용될 것을 주장하였다. Hackathorn [1988] 정보공학의 방법론과 틀을 Depth:(방법론-기술-틀)와 Breadth:(계획과분석-논리적설계-시스템구현) 라는 틀(*framework*) 속에서 비교, 분석하였다.

○SISF의 성공,실패의 문제점 파악 : Lederer & Sethi [1988, 1991]는 SISF수

행상의 잠재적 문제점 파악하기 위해 정보담당 경영층과, 최고경영층을 상대로 설문조사를 하고 이를 요인분석(*Factor Analysis*)와, 판별분석(*Discriminant Analysis*)을 통하여 분석하고 있다. Goodhue et. al. [1992]은 31개의 사례연구를 중심으로 현재의 방법론이 완벽한 것이 아님을 주장하고 있다. 현재의 많은 기업들이 SISF의 수행으로 부터, 많은 비용의 투자와 함께, 불확실한 잇점, 실행의 어려움 등의 곤경에 처해 있으며 현재 수행하고 있는 SISF에 대한 반성을 촉구하고 있다. 또한 Earl [1993]은 27개의 기업을 대상으로 SISF를 5개의 기준으로 분류하여 그 사용경험을 분석한 결과 조직적 접근법이 가장 효과적이라는 결론을 제시하였다.

이와같이 SISF에 관한 기존연구는 주로 개념적 *framework*의 제시, 설문조사를 통한 방법론간의 비교분석, 또는 SISF 방법론의 적용시 성공과 실패에 영향을 미치는 요인을 규명하는 것으로 나타났다. 반면에 특정 방법론의 기술적인 향상을 위한 연구는 그 필요성에도 불구하고 기술적인 어려움때문에 매우 부족한 실정이다. 방법론 개발에 관한 최근의 연구는 BSP의 개선을 위하여 Crockett et. al. [1989]이 군집분석방법을 활용하여 객관적인 체계를 통한 CASE환경을 제시하려고 시도하였으나 최종산출물의 해석상의 어려움때문에 활용상 그 한계가 있다고 할 수 있다. 기존 SISF 방법론을 보다 구체적으로 살펴보면 다음과 같다.

## 2.1 SISP의 제단계

정보시스템 전략적계획은 주요 서브시스템과 서브시스템간의 관계 정립을 통하여 기업 전체적인 관점에서 효과적인 정보시스템 구축을 위한 정보체계 (information architecture)를 제공한다. 정보시스템 전략적계획에서 산출된 서브시스템은 우선순위에 따라 시스템 분석 및 설계단계를 거쳐 활용가능한 시스템으로 구현되며 기업 전체 수준의 정보관리를 효율적으로 수행하기 위한 데이터베이스 구축의 기초가 된다. 따라서, 전략적 계획단계에서 기업에서 필요한 서브시스템을 제대로 파악하지 못하였을 경우에는 효율적인 시스템 구축이 이루어지지 못함에 따라 자원의 낭비와 효율적인 정보관리에 많은 지장을 초래하므로 서브시스템을 효율적으로 정확히 파악하는 과정이 정보시스템 전략적 계획에서는 매우 중요한 과제이다.

경영전략의 반영 뿐만아니라 서브시스템의 효율적 파악을 위하여 다양한 SISP 방법론들이 개발되었으며 방법론들 간에는 여러 차이점이 있을 수 있으나 SISP의 일반적인 과정을 살펴보면 다음의 4단계로 진행된다 [Brancheau et al, 1992].

① 조직의 전사적 모형을 개발-기능(functions), 프로세스(processes), 수행활동(activities)의 파악

② 정의된 프로세스 또는 수행활동들에 의해 이용되어지는 데이터베이스를 규명한다.

③ 유사성분석(affinity analysis)을 수행하여 밀접하게 연관된 데이터베이스들을 그룹핑한다.

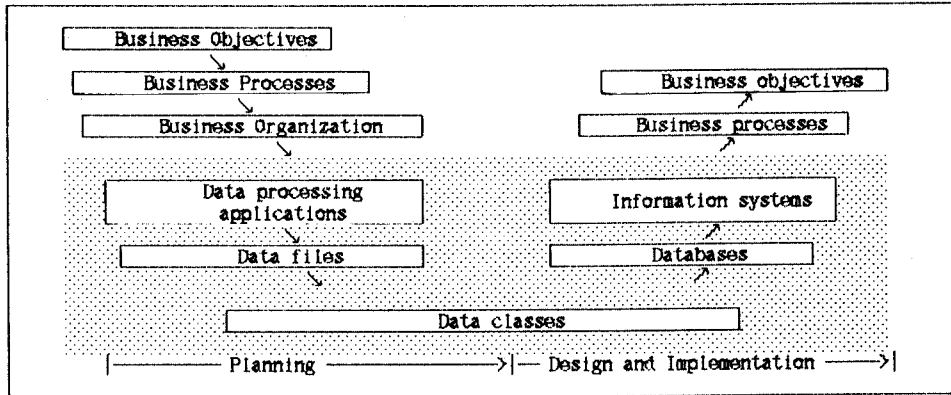
④ 데이터클래스와 이들을 이용 또는 생성하는 프로세스들을 그룹핑하여 구현을 위하여 서브시스템인 주요시스템 영역을 규정한다.

첫번째와 두번째 단계는 경영지향적 접근 방법으로서 주로 경영의 목표와 관련하여 정보시스템과는 무관하게 주요 기능별로 관련 프로세스와 데이터베이스를 규명하는 단계이므로 체계적 분석절차를 통한 자동화에는 별 효과가 없다고 할 수 있다.

반면에, 세번째와 네번째 단계는 데이터베이스의 그룹핑을 통한 데이터클래스간의 그룹형성, 이들 데이터클래스와 프로세스간의 그룹핑을 통한 서브시스템 규정 등 많은 절차가 따르는 과정을 거치므로 객관적이고도 체계적인 방법론 개발에 따른 생산성 향상이 요구되는 중요한 단계이다.

이와 같은 과정을 통하여 정보시스템 체계가 완성되며 이를 바탕으로 조직의 데이터클래스, 주요 응용영역, 현 시스템에서 바람직한 시스템으로의 이전계획, 목표 응용프로그램 개발을 위한 자원할당 권고사항 그리고 H/W, S/W와 데이터에 관한 표준의 판단근거가 제시된다.

SISP의 제단계를 지원하는 주요방법론으로는 서론에서 언급하였듯이 IBM의 Business Systems Planning(BSP), James Martin의 Information Engineering(IE), Strategic Systems Planning, Method/1,



〈그림 2-1〉 SBP방법의 제단계

Information Quality Analysis [Vacca, 1984], Critical Success Factors(CSF), Value Chain Analysis 등 다양한 방법론이 있다. 이 방법론 중에서 IBM에서 제시한 BSP 방법론이 경영지향적(business-oriented)인 접근방법으로서 하향식(top-down)계획과 상향식(bottom-up)구현 방법의 사용, 사용의 용이성, 체계적인 문서화를 지원함으로써 SISP 방법론 중에서도 가장 대표적인 방법론으로 널리 활용되고 있기에 [Lederer and Sethi, 1988] 본 논문에서는 BSP의 개선에 중점을 두고자 한다.

## 2.2. BSP방법

BSP방법은 제품수명주기에 따른 배열을 이용하여 기업경영을 지원하는데 필요한 논리적으로 연관된 데이터의 범주인 데이터클래스를 그룹핑함으로써 일명 수명주기순서 방법이라고도 한다 [IBM, 1981]. BSP의 각 단계를 살펴보면 〈그림 2-1〉과 같다.

BSP방법은 하향식 계획과 상향식 구현의 실행전략에 따라서 정의된 시스템들이 시간의 경과에 따라 모듈별 벽돌쌓기 방식으로 개발되며, 한편으로는 조직내의 업무우선순위, 사용가능한 자원 그리고 여타의 단기적 고려사항들과 보조를 맞추도록 하게 한다는 것이다. 전사적 정보시스템은 한 번의 프로젝트로 수행하기에는 너무 크고, 개별적으로 구축하기에는 개발 우선순위의 문제, 데이터의 불일치, 통합시스템 디자인의 불가능 등 여러문제점이 수반된다. 이러한 문제를 극복하기 위하여 먼저, 기업경영의 목적을 분석하고 목적을 달성하기 위한 프로세스와 관련 조직의 파악, 프로세스의 수행에 필요한 데이터를 확인하여 데이터클래스를 형성하는 하향식 분석을 하여 전사적 수준의 데이터베이스를 형성하고, 이러한 틀 속에서 개별적인 시스템을 구축하여 장기적으로 효율적인 정보시스템을 구축하는데 도움을 준다.

이같은 장점에도 불구하고, BSP방법은 단순순작방법에 의존하고 있으며, 경우에 따라

서는 서브시스템이 적절히 정의되지 못하는 사례도 발생할 뿐 아니라, 데이터클래스의 순서변화에 따른 서브시스템 확정시 전문가의 판단에 따라 복수해석이 발생하기도 하여 전문가에 따라 정보체계가 다르게 정의되는 문제점이 있다.

즉, 데이터클래스를 그룹핑하기 위해서는 먼저, 조직의 프로세스들이 수명주기적 순서로 배열되어야 한다. 일반적으로 서비스 및 생산조직들은 계획(planning), 획득(acquisition), 성숙(stewardship), 폐기(disposal) 라는 4단계의 수명주기를 가진다고 볼 수 있다. 데이터클래스를 열(column)로 세우고, 프로세스와의 교차지점에 데이터베이스가 프로세스에 의해 사용되면 'U'를, 생성되면 'C'를 부여한다.

데이터클래스의 순서는 'C'로 명명된 셀이 왼쪽 상단에서 오른쪽 하단으로 배열되도록 옮겨지게 된다. 그러나 프로세스의 순서는 고정시킨다. 종종 데이터클래스가 하나 이상의 프로세스에서 생성될 수 있으며, 이런 현상이 데이터클래스의 배치 순서결정을 어렵게 하고, 수행 전문가에 따라 상이한 배열순서가 이루어지기도 한다. 이러한 데이터클래스 순서유형의 차이가 결국 구현의 목적에 사용될 서브시스템 그룹핑 생성의 기초가 되므로 전문가에 따라 전혀 다른 결과가 제시되는 문제를 발생시킨다.

한편, 주요시스템영역(major system area)의 선정은 'C'의 대각정렬을 기본으로 하여 주관적으로 이루어는데 이 과정 역시

매우 개략적이며, 두 개의 서브시스템의 경계지역에 위치한 데이터클래스의 서브그룹선택에 관한 확실한 기준을 제시하지 못하고 있다. 또한 이들 주요시스템영역의 선정이 어느정도 정확하다고 하더라도 매트릭스에서 'C'의 대각정렬에 위배되는 'C'가 나타나는 문제가 발생함으로써 시스템 선정에 어려움을 가져다 준다.

이와같이 BSP방법은 정보체계내의 서브시스템을 구성하는 데이터클래스 그룹을 형성하기 위한 정형화된 기준이 없다. 데이터클래스간의 프로세스에 의한 생성(Create:C) 및 사용(Read:R)의 관계 정보에 기초하지 않고 단순히 프로세스의 배열순서에 따라 생성되는 데이터클래스를 재배치하고, 재배치에 따라 인접하게된 데이터클래스의 그룹을 하나의 서브시스템으로 파악하여 전체적인 정보체계를 형성하므로, 객관적인 기준을 가지지 못한다고 할 수 있다. 또한, BSP에서 데이터클래스는 하나의 프로세스에 의해 생성되기도 하고, 둘 이상의 프로세스에 의해 생성되는 데이터클래스도 있다. 그리고 이들 프로세스는 독자적 활동에 의해 데이터클래스를 생성하기도 하며, 여타의 이미 생성된 데이터클래스를 이용하여 프로세스가 수행되어져서 데이터클래스를 생성하기도 한다. 즉, 생성과 이용의 정보를 별개로 구분·분리하여 생성(C)이나 이용(R) 중 하나의 정보만으로 데이터클래스 간의 관계를 파악할 수는 없을 것이다. BSP방법은 생성(C)에 관한 정보만 이용하여 수명주기순으로 배

열린 프로세스의 위치에 의거한 데이터클래스의 재배치이므로, 모든 정보를 반영한다고 볼 수 없다.

따라서, 사용이 편리하며 해석이 용이하다는 BSP 방법론의 장점에도 불구하고 과도한 문서화에 따른 계획기간의 장기간 소요, 성공여부가 전문가에 자질에 의존, CASE tool 과 같은 컴퓨터 지원을 위한 체계적인 방법론 미비, 보다 많은 분석요구 등의 문제점이 지적되고 있는 실정이다.

### 2.3. 정준상관분석방법

이 방법은 Crockett et al. [1989]에 의해 BSP방법을 개선하고자 제시되었다. 정준상관분석(Canonical Correlation Analysis) 방법은 정준상관특점을 이용하여 군집을 형성하며, BSP방법과는 다르게 생성(Create: C)와 이용(Read: R) 모두를 고려하며, 행(프로세스)과 열(데이터클래스)의 위치이동을 통하여 군집을 형성한다. Crockett의 연구에서는 SAS의 PROC CANCORR의 결과에서 필요한 정보를 얻고 있다. 프로세스와 데이터클래스와의 관계를 나타낸 원래의 행렬에 대하여 정준상관분석을 실시한다. 정준변형특점(canonical score)을 계산하고, 인공변수인 정준변수를 기준으로 군집분석(PROC CLUSTER)을 수행하여 원래의 자료를 원하는 형태의 군집으로 형성한다.

정준상관분석방법은 통계적 방법론을 사용하여 서브시스템 생성과정을 객관화시키고

자동화가 가능한 프레임워크를 제시하였다는 점에서 이론적으로 기여가 있다고 할 수 있다. 하지만, 서브시스템을 생성하는 과정에서 경영목표에 따라 기능적으로 또한 논리적으로 정리된 프로세스가 정준상관분석과정을 통하여 데이터클래스와 그룹핑되는 과정에서 정리된 프로세스의 위치가 뒤바뀔에 따라 두 변수군간의 상관관계를 최대화할 수 있지만 이러한 상관관계에 의해 정렬되어 도출되는 서브시스템을 정의하는 데는 많은 어려움이 따른다.

〈그림 2-2〉는 정준상관분석의 결과에 설명의 편의를 위하여 형성된 서브시스템에 ①부터 ⑨까지의 번호를 부여하였다. 〈그림 2-2〉에서 보듯이 데이터클래스에 독립적인 책임을 가지는 서브시스템은 가장 오른쪽 아래쪽에 위치한 ⑨ 서브시스템 밖에 없다. 대부분의 프로세스는 데이터클래스 그룹을 불규칙하게 중복적으로 공유하고 있다.

정준상관분석방법을 따른 결과 〈그림 2-2〉에서 서브시스템의 중복에 따라 서브시스템을 정의하기도 어려울 뿐만아니라, 서브시스템의 개발우선순위의 파악 및 서브시스템간의 정보흐름을 파악하기가 매우 어려움을 알 수 있다.

이와같이 SISF에 있어서 서브시스템 생성과정은 일반적으로 비체계적이고 수작업에 따른 과다한 시간과 노력이 소요됨으로써 가장 중요한 과정중에 한 부분임에도 불구하고 개선을 위한 새로운 방법론의 개발이 부진한 실정이다. 특히 구조적 시스템 개발을 위해



Process	employees	financial	budget	planning	sales	territory	customer	order	production	work in progress	inventory	material	Fin. Goods	Inventory	Production	machines	tools
recruiting	R	R	R	R													
personnel planning	R	R	R	R													
compensation policy	R	R	R	R													
payroll	R	R	R	R													
funds management																	
capital acquisition																	
budget planning																	
financial planning																	
product range review																	
cash flow	R	R															
selling	R	R															
sales forecasting	R	R															
product pricing	R	R															
customer relations	R	R															
territory management	R	R															
sales administration	R	R															
creditors & debtors	R	R															
market analysis	R	R															
cost accounting																	
profitability analysis																	
order servicing																	
quality control																	
packing																	
product design																	
purchasing																	
inventory control																	
receiving																	
materials control																	
finished stock control																	
shipping																	
product spec. maint.																	
materials Requirements																	
plant scheduling																	
sizing and cutting																	
workflow layout																	
capacity planning																	
machine operations																	

<그림 2-2> 정준상관분석에 의한 서브시스템 도출결과

서는 다양한 CASE tool이 개발되어 있는 것과 비교할 때 서브시스템 생성과정을 지원하기 위해서 컴퓨터지원을 통한 체계적 접근 방법 개발에 관한 연구가 미진한 것은 매우 아쉽다고 할 수 있다.

따라서, 정준상관분석 방법의 단점을 개선하여 BSP 방법론의 해석상의 용이성을 제공함은 물론 통계적방법론 등을 사용한 객관적이고도 체계적인 방법론을 개발하여 SISP를

위한 CASE환경을 제공하는 것이 매우 중요한 연구과제라 할 수 있다.

### III. 컴퓨터지원 통계적 접근방법의 개발

BSP 방법론의 해석상의 용이성을 제공하기 위해서는 경영환경의 분석에서 파악된 경영 프로세스를 그대로 유지하면서 고정된 프로세스와 상관관계가 높은 데이터클래스를

그룹핑하는 것이 바람직하다. 이를 지원하는 통계적 방법론으로는 변수군집분석(Variable Clustering)방법이 타당한 것으로 파악되었으며 변수군집분석 방법을 통하여 그룹핑된 서브시스템을 보다 완벽한 형태로 구축하기 위해서 서브시스템 생성규칙을 개발하였다. 새로운 방법론을 통하여 정보체계 형성과정이 보다 체계화되고 컴퓨터지원이 가능한, 즉 CASE 환경지원이 가능한 장점을 제공한다고 할 수 있다.

### 3.1. 변수군집분석 방법

변수군집분석은 일종의 다중그룹요인분석(multiple group factor analysis)중의 사각성분분석(oblique component analysis)으로 변수들 간의 상관관계를 이용하여 상관성이 높은 변수들이 하나의 그룹이 되게 한다. 변수들의 그룹핑은 “집단으로 묶여진 변수들 간의 상호상관관계가 전체집합 내에 속해 있는 다른 변수들과의 상관관계 보다 높은 상관관계가 있다[Harman, 1976]라는 가정 하에 수행된다. 여러 개의 변수가 동일 성분으로 측정되는 공통성분을 가지면 이것이 하나의 그룹이 된다.

개별 그룹에 속하는 변수들 간에는 선형관계가 존재한다고 할 수 있으므로, 변수집단을 중복되지 않는 그룹으로 나누면, 이들 그룹은 각각이 일차원으로 해석될 수 있다. 군집성분(cluster component)을 계산한 후 이들 군집성분들의 분산을 이용하여 그룹간의

변화량(variation)의 합이 커지도록 하여 그룹을 분할한다.

변수군집분석은 어떤 현상을 설명하는데 필요한 50가지 항목을 보유하는 테스트자료가 있다고 할 때, 이들 항목을 관련된 테스트끼리 묶여진 5개의 군집으로 분류하는데 이용할 수 있으며, 분류된 각 군집은 하위테스트(sub-test)로 취급할 수 있다.

정보체계의 프로세스\*데이터클래스 행렬의 자료구조에서 데이터클래스는 변수로, 프로세스는 관측으로 해석된다. 변수군집분석은 일반적인 군집분석과는 달리 유사한 관측의 그룹을 목적으로 하지 않고 변수의 그룹을 그 목적으로 한다. 관측을 대상으로 하는 군집분석에서의 분류기준은 거리(distance)의 개념을 이용하며, 각 관측들의 거리를 계산하여 큰 거리값을 갖는 관측들을 분할된 그룹으로 분류한다. 이에 관하여 원래의 프로세스\*데이터클래스 행렬을 데이터클래스\*프로세스 행렬로 전치시키고, 그룹핑의 대상이 되는 데이터클래스를 관측으로 위치시킨 후, 일반적인 군집분석을 수행할 수도 있을 것이다. 그리고, 거리의 개념을 유사성(similarities)으로 환산하여 생각한다면, 유사성이 높은 관측들끼리 하나의 그룹으로 형성한다고도 볼 수 있다.

그러나, 서브시스템의 형성은 관측(전치된 변수)들 간의 거리의 개념으로 파악해서는 안된다. 데이터클래스의 그룹핑은 구현가능한 서브시스템을 형성하게 되어, 개별적 응용프로그램 또는 데이터베이스의 논리적 생

성·유지의 총괄적 책임을 가지는 것으로, 하나로 그룹핑되는 데이터클래스간에 공통의 이용목적을 가진다고 볼 수 있다. 이렇게 볼 때, 변수군집분석의 개념은 본 연구와 일치점을 가진다고 할 수 있다. 분석에서 군집성분이 공통성분의 역할을 하며, 이들 공통성분간의 분산이 커지도록 그룹핑을 하게 되고 이는 바로 개념적으로 공통의 목적에 이용되는 데이터클래스 그룹핑과 일치한다고 할 수 있다.

본 연구에서는 SAS의 PROC VARCLUS의 절차를 이용하여 변수군집을 수행한다. PROC VARCLUS 절차를 살펴보면 다음과 같이 3단계로 진행된다.

① 기본적으로 모든 변수를 하나의 그룹으로 보고 시작한다. 프로세스\* 데이터클래스의 상관관계행렬을 이용하여, 변수의 집합 변수인 합성변수(composite variable;  $T_p$ )를 구한다.

$$T_p = \sum(z_k; k \in G_p) \quad (p = 1, 2, \dots, m) \quad m$$

은 그룹의 수를 나타내고, z는 개별 변수들을 나타낸다. 그런 후 분할 대상 그룹을 선정하여 그룹을 둘로 분할하게 된다. 분할대상 군집의 기준은 군집성분에 의해 설명되어지는 변동이 작은 그룹을 선택한다.

단일 합성변수에서 분할된 합성변수  $T_p$ 들의 분산은  $ST_p^2 = \sum(rik; j, k \in G_p)$ 이고, 어떤 두 합성변수간의 상관관계는  $r_{T_p, T_q} = \sum T_p T_q / N ST_p ST_q$ 로 나타낸다( $r$ 은 상관관계). 즉,  $r_{T_p, T}$

이 가장 큰 합성변수들로 분할한다.

② 선택된 군집은 직교회전(orthoblique rotation)을 시켜서 제일 큰 값과 두번째로 큰 값의 주성분을 발견하고 이들을 이용하여 두 개의 군집으로 분할한다. 분할된 군집에 변수들을 할당할 때는 변수들 간의 결정계수( $r^2; w_{ip}$ )가 커지도록 할당한다.

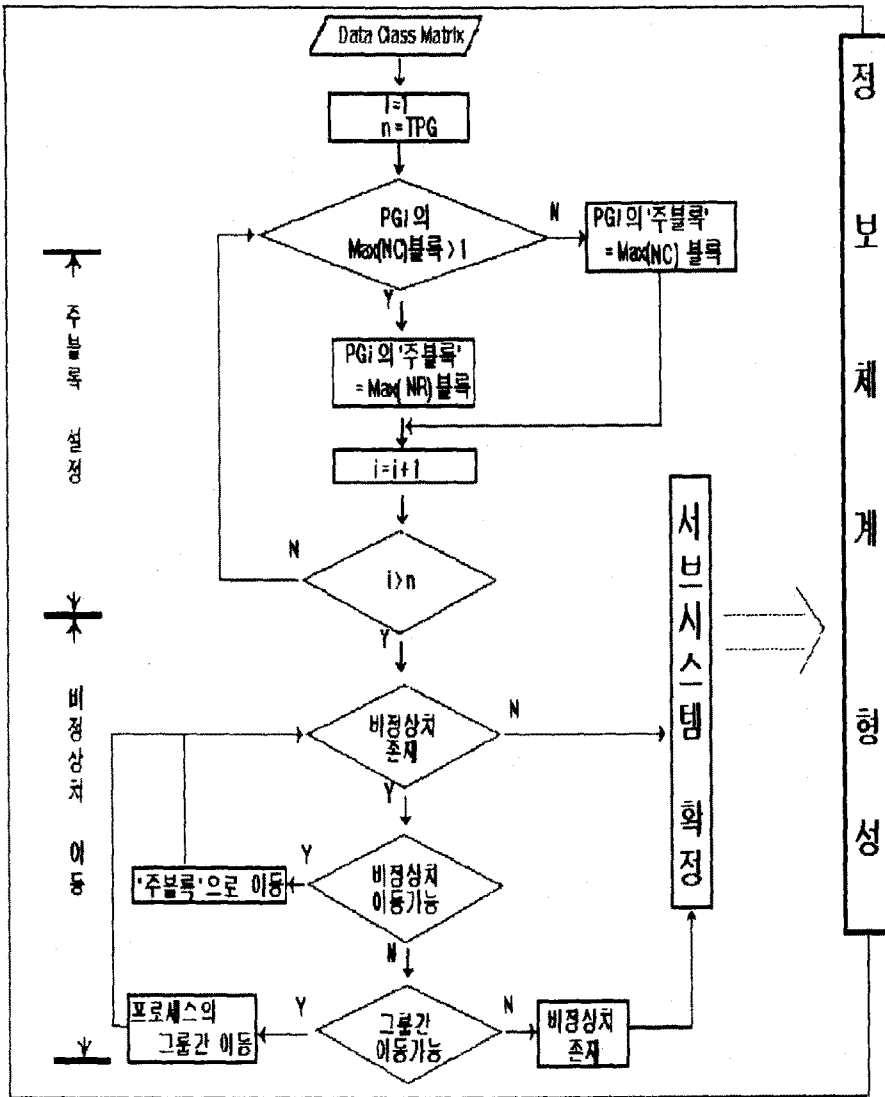
$$W_{ip} = \sum(r_{jk}; k \in G_p) \\ (j = 1, 2, \dots, n; p = 1, 2, \dots, m) \\ n \text{은 변수의 갯수}$$

③ 이러한 작업은 더 이상 분해할 대상기준을 만족하는 군집이 나오지 않을 때까지 계속 하게 된다.

### 3.2. 서브시스템 생성 규칙

데이터클래스는 변수군집분석을 통하여 다른 데이터클래스간 보다 높은 상관관계를 나타내는 데이터클래스를로 그룹핑(grouping)되어 하나의 서브시스템에서 생성·유지되는 데이터베이스를(database)를 형성하게 된다. 하지만 변수군집분석을 통한 데이터클래스 그룹핑은 완벽한 형태의 정보체계(Information Architecture)를 형성하지 못하고 있다. 변수군집분석의 결과가 완성된 정보체계에서 소수의 예외가 발생하고 있다.

변수군집분석에서 발생하는 소수의 예외는 정확한 서브시스템의 완성을 지연시키며, 이에 관해 전문가의 의견에 따른 주관적 해석으로 일관되고, 체계적인 서브시스템의 형성



〈그림 3-1〉 CASA-SISP방법중 서비스시스템 확정단계 순서도

◇ : 평가 □ : 처리

PG=Process Group. TPG=Total number, of PG.

NC=Number of 'C'. NR=Numberof 'R'

을 기대할 수 없게된다. 그러므로 이러한 소수의 예외를 서비스시스템에 효율적이며 체계적으로 포함시키는 적절한 조정 절차가 필요하며, 이것을 우리는 서비스시스템 생성 규칙이라 한다.

변수군집분석을 통한 데이터클래스 그룹의 정의와, 프로세스그룹을 바탕으로 행렬의 세로축인 프로세스그룹과 가로축인 데이터클래스그룹의 교차지점을 형성하고 이를 일종의 '블록'이라고 생각한다. 서비스시스템 생성에

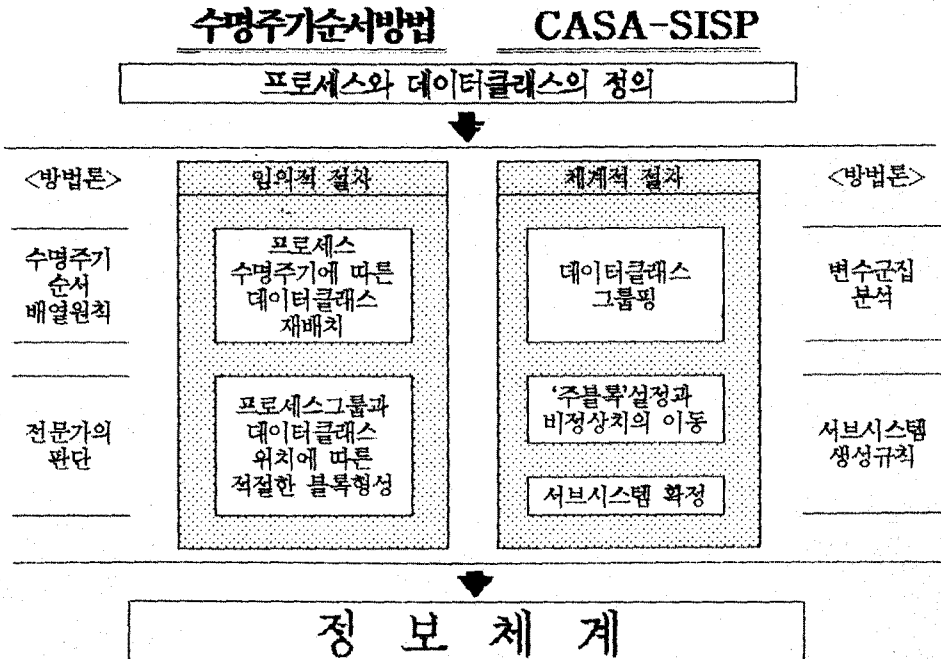
이용되는 생성(C)정보와 이용(R)정보 중 하나의 정보를 기준으로 하게 된다. 이때 이용(R)의 관계는 프로세스간 복잡하나, 생성(C)의 경우에는 데이터클래스가 프로세스에 의해 한번은 꼭 생성되므로 이를 중심으로 '서브시스템' 설계하는 것이 보다 단순하다. 기존의 BSP방법에서도 'C'정보를 서브시스템 설계의 기준으로 설정하고 있다. '블록' 중에서 우리가 관심을 두고자 하는 것은 'C' 또는 'R'이 포함된 블록이며 특히, 생성(C)에 관한 정보를 포함한 블록을 'C-블록'이라고 하고 서브시스템 설계를 위한 기초를 삼게 된다. 'C-블록'을 기초로 하여 서브시스템 형성의 기준이 되는 '주 블록'을 형성하고, '주 블록'에서 제외된 생성(C)정보는 비정상치로 분류하여, 이들 비정상치에 대한 이동규칙을 설정하여 '주 블록'에 포함시키고, 최종

적으로 서브시스템을 확정한다. 이때 '주 블록'은 프로세스 그룹간에 상호 중복되게 선정되어서는 안된다.

확정된 서브시스템간 데이터 흐름은 서브시스템에 포함되지 못한 'R-블록'을 이용하여 나타내고 서브시스템과 데이터 흐름으로 정보체계가 도출된다. 서브시스템 확정단계를 순서도로 표현하면 <그림 3-1>과 같다.

### 3.3. BSP 방법과의 비교

BSP방법과 CASA-SISP방법의 주요절차를 비교 정리하면 <그림 3-2>와 같다. 두 방법은 BSP의 체계를 따르므로, 데이터클래스와 프로세스의 정의에 관련된 내용은 동일한 정보를 이용한다. 정의된 데이터클래스와 프로세스로부터 정보체계를 생성하기 위하여



<그림 3-3> 수명주기순서방법과 CASA-SISP방법의 비교

두 방법은 각기 다른 절차를 가진다. 크게 이분하여 BSP방법을 임의적 절차라고 한다면, CASA-SISP방법은 체계적 절차라고 할 수 있다. BSP방법을 수명주기순의 배열원칙과 전문가의 판단에 따른 서브시스템 설계와 정보체계의 생성이라고 할 수 있다면, CASA-SISP방법은 통계적 기법을 이용하여 데이터클래스의 그룹을 확인한 후에, 이들을 기초로 하여 주블록과 비정상치 등의 개념을 설정하여 서브시스템을 확정하고 최종 정보체계를 수립하고 있다.

## IV. CASA-SISP의 적용

### 4.1. BSP방법의 적용

〈그림 4-1.A〉의 데이터 행렬은 본 연구에 이용된 원시 데이터이며, 〈그림 4-1.B〉의 데이터 행렬은 BSP방법에 따라 도출된 최종 정보체계이다. BSP방법에서 원시 데이터에 적용된 기법은 전문가의 검사에 따라, C를 대각선 방향으로 형성되도록 데이터클래스를 이동하여 정보체계를 완성하게 된 것이다. CASA-SISP 방법론을 〈그림 4-1.A〉 데이터 행렬을 사례로 적용함으로써 CASA-SISP 방법론이 BSP방법론과 동일한 결과를 제공하는지를 살펴보고자 한다.

### 4.2. 변수군집분석과 데이터클래스 그룹핑

입력 데이터인 프로세스 \* 데이터클래스의

자료행렬은 'C'와 'R'이라는 명목척도로 구성되어 있으며, 이들은 분석을 위하여 구간척도로 변환되어질 필요가 있다. 구간척도로의 변환에서 BSP방법은 생성에 대한 정보에만 의미를 두어서 생성(C)은 1로 이용(R)은 0의 값을 부여한 형태로 되었다. 정준상관분석방법에서는 생성(C)과 이용(R)을 동일한 1로 변환했다.

본 연구에서는 적절한 가중치의 차를 미리 알 수 없으므로 가중치 차를 1 부터 8까지 변화시켜가며 적절한 수준을 확인하고자 한다. 적절한 수준의 가중치 차는 서브시스템 설계에 필요한 오퍼레이션의 수를 감소시킬 수 있으며, 이는 보다 정확한 서브시스템을 얻는데 효과적이라고 할 수 있다. 많은 시뮬레이션을 거친결과 가중치의 차가 2인 (즉, 'C'=3, 'R'=1) 경우가 가장 정확한 결과를 제시하였다. 즉, SAS의 PROC VARCLUS를 통하여 분석한 결과 〈그림 4-2〉와 같이 프로세스와 관련 데이터클래스의 그룹핑이 형성되었다.

즉, 프로세스 \* 데이터클래스 행렬의 자료구조에서 데이터클래스는 변수로, 프로세스는 관측으로 해석된다. 변수군집분석은 일반적인 군집분석과는 달리 유사한 관측의 그룹을 목적으로 하지 않고 변수의 그룹을 그 목적으로 한다. 본 연구가 추구하는 해석의 용이성을 위해, 최종적으로 산출되는 정보체계의 해석 기준인 프로세스를 고정하기 위해서는 변수군집분석이 효과적이며, 일반적인 군집분석이 관측값들 간의 거리 개념을 이용하

DATA CLASS		W				
market analysis	R		RR	R		R
product range review					RR	R
sales forecasting	RC		RR	C		R
financial planning	R		R			
capital acquisition	RC		R			
funds management	RR		R			
product design					CCC	
product pricing	R		R		C	CC
product spec. maint.					C	RR
materials Requirements						RR
purchasing	OC	RR				
receiving	ORR					R
inventory control		C			R	
quality control						
capacity planning	R	C				
plant scheduling				RC	RR	
workflow layout				C	RR	
material control	R					
mixing and cutting						RR
machine operations						R
facility management	C		R		CC	
selling				RR		
sales administration	R					
customer relations						
finished stock control			C			RR
order servicing	R		C			RR
packing			R			RR
shipping						
conditions & debtors	R		RR		RR	RR
cash flow					CC	RR
general						RR
cost accounting	R			C		R
budget planning	CR			RR		
profitability analysis						RR
personal planning	R				CC	
recruiting						RR
competition policy	R					RR

(그림 4-1.A) 초기 데이터 행렬

DATA CLASS		W				
market analysis	RR		RRR	R		R
product range review	R				RRR	R
sales forecasting	RC		RR	C		R
financial planning	R		R			
capital acquisition	RC		R			
funds management	RR		R			
product design	R				CCC	
product pricing	R		R		C	CC
product spec. maint.					C	RR
materials Requirements						RR
purchasing	OC	RR				
receiving	ORR					R
inventory control		C			R	
quality control						
capacity planning	R	C				
plant scheduling				RC	RR	
workflow layout				C	RR	
material control	R					
mixing and cutting						RR
machine operations						R
facility management	C		R		CC	
selling				RR		
sales administration	R					
customer relations						
finished stock control			C			RR
order servicing	R		C			RR
packing			R			RR
shipping						
conditions & debtors	R		RR		RR	RR
cash flow					CC	RR
general						RR
cost accounting	R			C		R
budget planning	CR			RR		
profitability analysis						RR
personal planning	R				CC	
recruiting						RR
competition policy	R					RR

(그림 4-1.B) 도출된 정보체계

는 것 과는 달리 공통성분을 가지는 것으로 추측되는 변수들은 하나의 군집으로 형성한다.

4.3. 서브시스템 확정

서브시스템은 데이터 흐름과 함께 정보체계를 형성하는 주요구성부분으로서 데이터클래스 그룹과 프로세스 그룹의 관계를 블록으로 묶어서 프로세스 그룹 이름을 부여한 것이다. 그러므로 BSP와 같은 계획단계에서 생성되는 서브시스템은 분석단계에서 세부적으로 다루어지는 영역(Business Analysis

Areas)을 나타낸다 하겠다. 서브시스템 확정의 단계는 '주 블록' 확인과 비정상치 이동의 두 과정으로 구성된다.

1) '주 블록' 확인

(그림 4-2)는 가중치의 차를  $2('C'=3, 'R'=1)$ 로 하고, 그룹의 수를 15로 하는 경우의 데이터클래스 그룹핑 결과 그림이다. 세로축은 프로세스를 나타내고, 가로축은 데이터클래스를 나타낸다. 세로축을 구분하는 가로선은 프로세스 정의에서 파악된 프로세스 그룹을 의미하고, 가로축의 점선 세로선은 데이터클래스의 그룹핑을 의미한다. 가로

Process	Data Class															
	customers	budget	financial	procurement	Inventory	Fin. Goods	payments	sales Territory	employees	sales facilities	shop floor	work in process	open	product	parts	bill of materials
market analysis	R							RR						R		
product range review	R						R	RR						R		
sales forecasting																
financial planning							R									
capital acquisition																
funds management																
product design			R													
product pricing		R					R									
product spec. maint.																
materials Requirements																
purchasing																
receiving																
inventory control																
quality control																
capacity planning					R											
plant scheduling																
workflow layout																
materials control					R											
sizing and cutting																
machine operations																
territory management																
selling																
sales administration																
customer relations																
finished stock control																
order servicing																
packing																
shipping																
creditors & debtors																
cash flow																
payroll																
cost accounting																
budget planning																
profitability analysis																
personnel planning																
recruiting																
compensation policy																

〈그림 4-2〉 '주블록'확정(기준치 차 2('C'=3 'R'=1), 그룹의 수=15)

주: (■, ▨: 비정상치)

선과 세로선에 의해 구분되어진 사각형이 블록이며, 음영으로 표시된 부분은 'C-블록'을 나타낸다.

〈그림 4-2〉에서 'C-블록'의 수는 17개로 개별 블록을 기준으로 판단한다. 'C-블록'을

과약한 후에는 '주 블록'을 확인하게 된다.

〈그림 4-2〉에서 첫번째 프로세스 그룹의 경우, 두번째 데이터클래스 그룹과 3번째 데이터클래스 그룹의 교차점이 'C-블록'이고 이중 포함하는 생성(C) 정보가 가장 큰 'C-블



Process	financial budget	procurement	materials inventory	Fin. Goods Inventory	orders	costs	payments	sales territory	customer employees	shop floor facilities	machines	product design	bill of materials
market analysis	R							R	R	R			R
product range review	R							R	R	R			R
sales forecasting	C							R	R	R			R
financial planning	R							R					
capital acquisition	R												
funds management	R												
product design	R											C	C
product pricing	R							R				C	C
product spec. maint.												C	C
materials Requirements			R	C									
purchasing			R	C									R
receiving			R	R									R
inventory control			R	R							R		
quality control			R	C									
capacity planning		R		R						R	R	R	R
plant scheduling										R	R	R	R
workflow layout										R	R	R	R
materials control		R								R	R	R	R
sizing and cutting										R	R	R	R
machine operations										R	R	R	R
territory management					R			R	C				
selling					R			R	C				
sales administration					R			R	C				
customer relations					R			R	C				
finished stock control					C					R		R	R
order servicing					C					R		R	R
packing					C					R		R	R
shipping					R					R		R	R
creditors & debtors		R	R			R	R	R	R	R	R		
cash flow		R	R			R	R	R	R	R	R		
payroll						R	R	R	R	R	R		
cost accounting			R			C	C	C	C	C	C		R
budget planning	C	R	R			R	R	R	R	R	R		R
profitability analysis						R	R	R	R	R	R		R
personnel planning	R									C	C		
recruiting										R	R		
compensation policy	R									R	R		

〈그림 4-3〉 서브시스템 확정(가중치의 차=2, 그룹의 수=15)

록'은 두번째 데이터클래스이므로 '주 블록'으로 설정된다. 이때 '주블록'으로 선정된 'C 블록'과 비로 인접한 'C-블록'이 존재할 때 인접한 'C-블록'을 이미 선정된 '주블록'에 포함시켜 '주블록'을 결정한다. 예를 들면 financial 데이터클래스는 이미 주블록으로 설

정된 budget과 planning블록에 포함되어 주블록으로 최종적으로 결정된다. 〈그림 4-2에서는 이와같은 주블록 선정과정에 따라 프로세스 그룹수와 동일한 8개의 주블록이 결정되며 이 과정에서 선정이 안된 4개의 비정상치 블록이 존재하게 된다. 4개의 비정상치

는 다음과 같은 이동규칙에 따라 주블록으로 포함됨에 따라 서브시스템이 완성된게 된다.

## 2) 비정상치의 이동

비정상치의 이동은 비정상치 정보를 나타내는 프로세스 그룹 이외의 프로세스 그룹에서 생성정보를 나타내지 않는 경우에 한하여 '주 블록'안으로 이동할 수 있다는 규칙에 따른다.

〈그림 4-2〉에서는 첫번째 프로세스 그룹의 financial 데이터클래스, 두번째 프로세스 그룹의 bill of materials 데이터클래스 등이 '주 블록'을 형성하지 못하는 비정상치로 분류된다. 세번째 프로세스 그룹의 비정상치 중 하나인 materials inventory 데이터클래스의 경우는 다른 프로세스 그룹에 생성(C) 정보를 가지지 않으므로 '주 블록'으로 이동시킬 수 있다. 그러나 일곱번째 프로세스 그룹의 비정상치인 budget 데이터클래스는 첫번째 프로세스 그룹에서 생성(C) 정보를 가지므로 만약 일곱번째 프로세스 그룹의 '주 블록'으로 이동하게 되면, 첫번째 프로세스 그룹의 '주 블록'이 파괴되고 비정상치가 발생되므로 '주 블록'으로 이동할 수 없다. 이러한 경우는 예외의 경우로 처리되며, 분석단계에서 세부적으로 처리된다.

예외의 경우는 BSP방법에서도 존재하고 있으며, BSP의 수명주기 체계를 따르게 될 때 불가피 하게 발생하게 된다. 이 문제의 극복은 예외로 처리된 생성(C)정보의 프로세스에 더 이상의 생성(C)정보가 없다면, 비

정상치 이동에 제약이 되는 프로세스 그룹으로 프로세스를 이동시켜서 재 배치 하는 방법이 가능하다 하겠다. 〈그림 4-3〉은 비정상치 이동규칙에 따라 비정상치를 재배치하여 서브시스템의 확정된 것을 나타낸다.

모든 프로세스 그룹의 비정상치 이동이 완료되면 서브시스템이 확정되고 정보체계를 형성한다. 서브시스템간의 데이터흐름을 나타낸 최종적인 정보체계는 부록의 〈그림 부. 4-1〉에서 보여주고 있다. 이러한 정보체계의 형성은 BSP방법과 동일하며, 서브시스템에 포함되지 못한 이용(R)정보를 이용하여 서브시스템간의 데이터흐름을 표현한다.

## V. 실험 및 결과분석

### 5.1. 민감도 분석

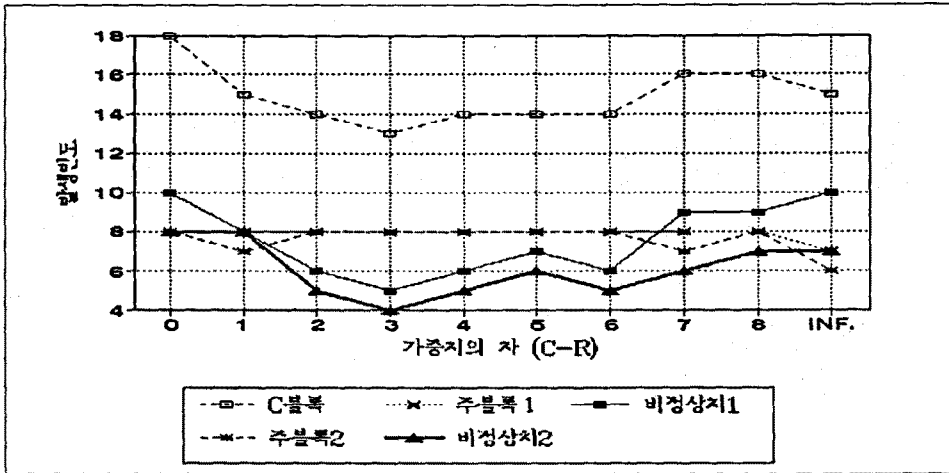
상관관계분석에서 생성과 이용의 정보를 명목척도에서 구간척도로 변환할 때, 가중치의 차를 두었다고 하였다. 그러나 구체적인 가중치의 차를 알 수 없으므로 이를 여러 값으로 변화시키고, 데이터클래스 그룹핑에서 그룹핑되는 그룹의 수가 그룹내에 포함되는 데이터클래스들 간의 동질성 여부에 영향을 주므로, 그룹의 수를 변화시켜서 그 민감도를 비정상치의 수로써 나타내고자 한다.

민감도 분석을 위하여 다음의 가정을 하였다. 가정 1: 생성(C) 과 이용(R)의 정보간에는 정보의 중요도에 차이가 있다.

가정 2: 그룹의 수가 많아짐에 따라 비정상

〈표 5-1〉 생성과 이용정보의 가중치 차의 변화

C	1	2	3	4	5	6	7	8	9	1
R	1	1	1	1	1	1	1	1	1	0
가중치의 차	0	1	2	3	4	5	6	7	8	$\infty$



〈그림 5-1〉 가중치 차에 따른 비정상치(outlier)의 변화(그룹 수 = 15)

치는 줄어든다.

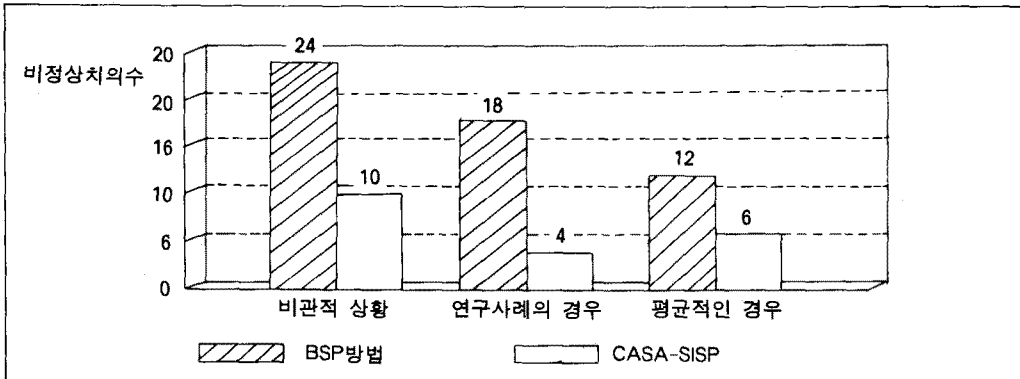
생성과 이용의 정보를 BSP방법의 경우에는 극단적으로 차별하고, 정준상관분석방법에서는 동일하게 다루므로, 본 연구에서는 어느 정도의 차이가 있을 것으로 보고 있다. 두 번째 가정은 정확한 그룹의 형성을 위해서는 그룹의 수가 많아져야 할 것이고, 그룹이 잘 형성되면 비정상치의 발생도 상당히 줄어들 것이라는 기대를 내포하고 있다.

민감도 분석을 위해 그룹의 수는 8, 10, 12, 15로 각각 변화시켰고, 그룹의 수 각 경우에 가중치의 차를 0에서 8까지로 분석을 하였다.

〈그림 5-1〉은 그룹의 수를 15개로 하였을 경우의 가중치에 따른 'C-블록', '주 블록',

비정상치(outlier)의 수를 보여주고 있다. 〈그림 5-1〉에서 '주 블록'과 비정상치가 각각 두 가지씩 있음을 알 수 있다. '주 블록'을 형성할 때 '주 블록'을 선정하고 그 인접한 'C-블록'을 '주 블록'내에 포함 하느냐 않느냐의 규칙에 따른 비정상치 결과를 보여주고 있다. 'X'와 '■' 기호는 포함하지 않는 규칙(규칙 1)을, '\*'와 '▲'는 포함하는 규칙(규칙 2)을 따른 경우이다.

규칙 1을 따를 경우에는 10개의 그룹수에서 가중치의 차를 3으로 했을 때(C=4, R=1) 비정상치의 수가 5개로 가장 작게 나타남을 알 수 있으며, 규칙 2를 따를 경우에는 그룹의 수 10과 15에서 비정상치의 수가 4개로 가장 작았다.



〈그림 5-2〉 BSP방법과 CASA-SISP방법의 비정상치(outlier)의 수 비교

민감도 분석의 결과는 가정 1을 만족하지 만, 가정 2의 경우는 ‘주 블록’ 형성 규칙에 따라 규칙 2의 경우에는 가정을 만족하였으나, 규칙 1을 따를 경우에는 데이터클래스 그룹의 수를 많게 했을 때, 비정상치의 수도 증가하여 가정을 위반하였다. 본 연구에서는 민감도 분석의 결과 중 ‘주 블록’ 형성은 규칙 2를 따르고, 그룹의 수 15에서 가중치 차 2에서 개선방법의 기준을 설정하고 있다. 이는 그룹의 수가 데이터클래스 수의 절반을 넘게 되면(사례의 경우 12개) 가중치의 차 변화에 덜 민감하게 반응하며, 민감도 분석의 결과 그림의 패턴에 의해 가중치의 차 2에서 수렴함을 근거로 하고 있다.

## 5.2. BSP방법과의 비교

비정상치(outlier)의 측면에서 BSP방법과 개선방법을 비교하여 보자. 개선방법에서 비

정상치는 개별프로세스그룹에서 ‘주 블록’에 포함되지 못한 생성(C)정보를 가지는 데이터클래스로, 결국에는 이동 규칙에 따라 ‘주 블록’에 포함되게 된다.

이러한 비정상치(outlier)의 개념을 BSP방법에 적용한다면, 하나의 프로세스에서 두 개 이상의 데이터클래스가 인접하고 있다면 이들을 한번에 이동시킬 수 있을 것이고, 그렇지 못한 경우에는 모두 ‘한번의 이동’이라는 작업을 거쳐야 하며 이들을 모두 비정상치라고 할 수 있다. BSP방법을 원래의 행렬과 동일한 빈 행렬을 바로 옆에 마련해 두고, 프로세스 수명주기순서에 따라 생성되는 데이터클래스를 빈 행렬의 왼쪽 우선순위로 배열한다고 할 때, 가장 낙관적인 상황에서는 한번에 모든 걸 옮길 수 있을 것이고(이렇게 되면 실제로 안 옮겨도 됨), 가장 비관적인 상황에서는 데이터클래스의 수 만큼의 이동회수가 필요하게 될 수도 있다.

CASA-SISP방법에서는 민감도 분석의 결과로 볼 때, 4번 이하의 이동만 하면 서브시스템이 확정된다.

BSP방법의 경우 비관적 상황과 낙관적 상황을 동일한 비중으로 고려할 때, 비관적 상황을 데이터클래스 수 만큼의 이동이 필요한 것으로 보아, 본 연구 사례의 경우 24번의 이동이 발생하고, 낙관적 상황은 이동의 필요가 없는 경우로 데이터클래스의 이동이 발생하지 않는다. 이러한 가정하에서 비정상치(outlier)의 단순평균 값은  $(0+24)/2$ 로 12가 될 것이다.

제시된 프로세스\* 데이터클래스 행렬 경우에서는 앞서의 방식대로 하면 18번의 이동이 발생하게 된다. 이를 CASA-SISP방법과 비교한 것이 <그림 5-2>이다.

CASA-SISP방법의 경우 각 상황은 민감도 분석의 실험 결과에서의 비관적, 본 사례 연구, 평균적 경우로 나타내었다. 특히, 평균적 경우는 실험된 모든 경우의 비정상치 수  $(238)/\text{실험회수}(40)$ 로 나눈 값이다.

### 5.3. 개선방법의 의의

CASA-SISP방법은 정보체계를 구성하는 서브시스템 설정을 위한 체계적 방법을 제공한다 하겠다. CASA-SISP방법은 변수군집 분석과 상관관계분석 등의 통계적 기법을 적용하여 데이터클래스를 그룹핑하고 그룹핑된 데이터클래스 정보를 바탕으로 체계적인 규칙을 적용하여 서브시스템을 확정하였다. 또

한 CASA-SISP방법은 통계적 기법을 사용한 정준상관분석 방법의 단점인 해석에 관한 문제를 해결하기 위하여 프로세스의 위치를 변경시키지 않음으로 해서 정보체계를 쉽게 형성할 수 있다.

통계적 기법의 적용을 통한 서브시스템 설계 기준의 설정을 CASA-SISP방법의 의이라 할 수 있으며, 이러한 기준의 설정에 따라 서브시스템 형성의 절차가 평균적으로 50% 절감됨을 알 수 있다.

본 연구가 제시한 CASA-SISP방법의 한계점 및 향후과제는 서브시스템 확정을 위한 '주 블록' 설정에서 규칙 1과 규칙 2를 적용했을 때, 연구사례의 경우에는 규칙2의 경우가 보다 나은 결과를 제시하였으나, 서브시스템 생성규칙의 다양한 사례 연구를 통한 검증이 필요하다 하겠다.

비정상치의 수를 최소로 할 수 있는 적정 그룹의 수를 선정함에 있어 체계적인 방법의 모색이 고려되어야 할 것이다. 그리고, CASA-SISP방법의 전 과정을 일반적인 상황에 대처하도록 프로그래밍하여, 최종정보체계 산출과정을 자동화하는 시스템의 개발이 필요하다 하겠다.

## VI. 결 론

BSP의 자동화라는 측면에서 본 연구는 BSP의 체계내에서 가장 중요한 문제라고 생각되는 서브시스템설계에 대하여 통계적기법을 이용한 방법을 제시하고 있다.

기존의 BSP방법에서는 수작업에 의존하고, 전문가의 판단에 따라 다른 해를 발생시킬 수 있는 점과, 프로세스의 데이터클래스 사용에 관한 정보는 전혀 고려되지 않는 점 등의 문제점을 인식할 수 있다. BSP방법의 대안으로 제시된 정준상관분석방법의 인공변수의 도입에 따른 해석상의 난점과 서브시스템 설계를 위한 기준의 상실등의 문제점을 파악할 수 있다.

본 연구는 BSP방법과 정준상관분석의 단점을 극복하기 위하여, 두 방법의 장점을 유지하고, 변수간의 상관관계를 이용하여 공통성분에 대한 상관관계가 커지는, 즉 상이한 그룹간의 변화량이 증가하는 변수군집의 가정을 기반으로 변수군집분석과 서브시스템 생성 규칙을 적용한 CASA-SISP방법론을 제시하였다.

본 연구는 기존의 방법론에 대하여, 프로세스에 의한 데이터클래스의 생성(C)과 이용(R)에 관한 정보를 모두 고려하고 있으며, 군집분석기법을 사용하여 데이터클래스 그룹을 생성하여 서브시스템설계를 위한 기준을 제시하고 프로세스 그룹별 데이터클래스 그룹을 확정하기 위하여 체계적인 생성규칙을 설정하였다. 이에 따라 컴퓨터에 의한 자동화의 실현이 가능하여 서브시스템 설계에 따르는 노력을 줄이고 프로세스 및 데이터클래스의 분석 및 검증 등에 보다 많은 노력을 경주할 수 있도록 하고, 전문가 의존적인 서브시스템 설계를 보다 객관적인 기준하

에서 수행되도록 하며, 신속하고 정확한 정보체계의 생성이 가능한 장점이 제공된다고 할 수 있다.

본 연구의 한계 및 향후 연구과제는 다음과 같다.

첫째, 서브시스템 생성규칙에 대한 다양한 사례 연구를 통한 검증이 필요하다 하겠다.

둘째, 비정상치의 수를 최소로 하는 적정 그룹의 수가 존재하는지, 그룹 수의 변화와 그에 따른 비정상치(outlier) 개수 변화가 통계적으로 의미가 있는지 등에 관한 체계적인 조사가 필요하다고 생각된다.

이러한 문제점과 함께 본 연구의 향후 과제로는 다양한 기본자료에 대하여 연구방법을 적용하여 'C'의 변환값에 대한 민감도 분석과, BSP 연구의 최종결과물인 정보체계의 형성을 자동화하는 CASE툴 개발이 있으며, 형성된 정보체계의 타당성을 평가할 수 있는 기준을 찾는 것도 중요한 연구과제라고 할 수 있다. 또한 기업에서 Business Process Reengineering (BPR)기법을 도입하여 경영혁신을 활발히 전개하는 추세에서 SISP에서 구축된 정보체계와 BPR이 상호보완되어 현 시스템을 고려하여 구축된 정보체계가 BPR의 추진에도 도움을 줄 수 있도록 연계방법론이 필요함을 지적하고 있다 [Teng, Kettinger, Guha, 1992]. 따라서, 정보체계와 BPR을 연계할 수 있는 framework과 구체적인 방법론의 개발이 향후 중요한 연구과제라 할 수 있다.

## 참 고 문 헌

- 윤관중, [경영정보시스템(MIS) 기본계획수립 방법론 연구 (BSP를 중심으로)], 서울, 한국외국어대학교, 석사학위논문, 1988.
- 양연수, [MIS Planning 방법론 연구 (business systems planning을 중심으로)], 서울, 한국외국어대학교, 석사학위논문, 1987.
- 신동욱, [경영정보 시스템의 기본계획 수립방안에 관한 연구] 서울, 동국대학교, 석사학위논문, 1986.
- 김기영·전명식 『SAS 군집분석』, 서울;자유아카데미, 1989.
- 김기영·문권순·전명식, 『SAS 상관분석』, 서울; 자유아카데미, 1992.
- 우치수역 『데이터베이스 환경의 실현 및 관리』, 서울; 교학사, 1990. James Martin 원저.
- 한재민 문태수, “전략정보시스템 계획수립을 위한 상황모델”, 경영정보학 연구 제 3 권 제 1 호, 1993. 6.
- Aronson, J.E. and Klein, G.,(1989) “A Clustering Algorithm for Computer-Assisted Process Organization,” Decision Science V.20. 730-745.
- Boynton, A. C. and Zmud, R. W.,(1987) “Information Technology Planning in the 1990’s: Directions for Practice and Research,” MIS Quarterly, March pp.59-71.
- Brancheau, J.C. and Wetherbe, J.C.,(1987) “Key Issues in Information Systems Management,” MIS Quarterly(11:1), March, pp.34-45.
- Crockett, H.D., Slinkman, C. W. and Eakin, M. E.,(1989) “A Strategy for Database Planning,” Decision Science Institutes Conference.
- Duran, B. S., and Odell, P.L.(1974) Cluster Analysis:A survey. New York:Springer-Verlag.
- Earl, M. J.,(1993) “Experiences in Strategic Information Systems Planning,” MIS Quarterly Vol.17, No.1, pp.1-24.
- Finkelstein, C.(1989) Information Engineering-From Strategic Planning to Information Systems, Addison-Wesley.
- Goldsmith, N.,(1991) “Linking IT Planning to Business Strategy,” Long Range Planning, Vol. 24, No.6, pp.67-77.
- Goodhue, D.L., Kirsch, L.K., Quillard, J. A. and Wybo, M. D.,(1992) “Strategic Data Planning: Lessons From the Field,” MIS Quarterly,

March, pp.11-34.

Hackathorn, R.D. and Karimi, J.(1988) "A Framework for Comparing Information Engineering Methods," MIS Quarterly(12:2), June, pp.203-220.Hansen, G.W. and Hansen, J.V., (1992) Database Management and Design, Prentice Hall.

Harman, H. Harry,(1976) Modern Factor Analysis, 3rd, The University of Chicago Press.

IBM ,(1981) Business Systems Planning, IBM Manual #GE20-0527-3, July.

King, W. R.,(1978)"Strategic Planning for Management Information Systems," MIS Quarterly, 2(2), June , pp.27-37.

King, W. R.,(1988) "Evaluating an Information systems Planning Process," Long Range Planning, 21(5), October, pp.103-112.

Lederer, A. L. and Mendelow, A. L.,(1988) "Information systems Planning: Top Management Takes Control," Business Horizons 31. May-June, pp.73-78.

Lederer, A. L. and Sethi, V.,(1991) "Critical Dimensions of Strategic Information Systems Planning," Decision Sciences (22:1), Winter, pp.104-119.

Lederer,(1988) "The Implementation of Strategic Information Systems Planning Methodologies," MIS Quarterly (12:3), September , pp. 445-461.

Martin, J.,(1990) Information Engineering, BOOK I, Introduction, Prentice-Hall, Englewood Cliffs, NJ."

Martin,(1990) Information Engineering, BOOK II, Planning and Analysis, Prentice-Hall, Englewood Cliffs, NJ.

Nolan, R. L.(1979) "Managing the Crises in Data Processing," Harvard Business Review, Volume 57, Number 2, March-April, pp.115-126.

Premkumar, G. and King, W. R.,(1991) "Assessing Strategic Information Systems Planning," Long Range Planning, Vol.24, No.5, pp. 41-58.

Pyburn, P. J.(1983) "Linking the MIS Plan with Corporate Strategy: An Exploratory Study," MIS Quarterly(7:2), June, pp.1-14.

Sullivan, C. H. Jr.,(1985) "Systems Planning in the Information Age," Sloan Management Review (26:2), Winter, pp.3-12.

Teng, J. T. C, Kettinger, W. J., and Guha, "Business Redesign and Information Architec-



ture: Establishing the Missing Links," Proceedings of the International Conference on Informations Systems, 1992, pp. 81-89.

Zachman, J. A.,(1982) "Business Systems Planning and Business Information Control Study: A Comparison," IBM Systems Journal, Volume 21, Number 1, pp.31-53.

### ◇ 저자소개 ◇



공동저자 김진수는 연세대 상경대학 응용통계학과 학사, U. of Texas (Arlington) 경영정보학 석사, Louisiana State University에서 경영정보학으로 경영학박사를 취득하고 부산대학교 경영학과 조교수를 거쳐 현재 중앙대학교 사회과학대학 경영학과 조교수로 재직하고 있다. 주요 관심분야는 정보공학(Information Engineering), 정보시스템 전략적계획 방법론 개발, 개념적 DB 설계, 객 체지향 DSS, 중소기업 정보화 및 CIM 등이다.



공동저자 황철언은 부산대학교 경영학과를 졸업하고, 동 대학 대학원에서 석사학위를 받았다. 현재는 인제대학교 경영학과 조교로 근무하고 있다. 관심분야는 정보시스템계획 방법론과 전사적 DB 분석 등이다.

부록. CASA-SISP 방법의 결과

<p style="text-align: center;">D a t a  C l a s s</p> <p style="text-align: center;">P R O C E S S</p>	<p style="text-align: center;">p l a n n i n g</p>	<p style="text-align: center;">m a t e r i a l s</p>	<p style="text-align: center;">F i n a n c i a l</p>	<p style="text-align: center;">G o o d s</p>	<p style="text-align: center;">p a y m e n t s</p>	<p style="text-align: center;">s a l e s</p>	<p style="text-align: center;">c u s t o m e r s</p>	<p style="text-align: center;">s h o p i n g</p>	<p style="text-align: center;">b i l l i n g</p>
<p>market analysis product range review sales forecasting financial planning capital acquisition funds management budget planning product design product pricing product spec. maint. materials Requirements purchasing receiving inventory control quality control capacity planning plant scheduling workflow layout materials control sizing and cutting machine operations territory management selling sales administration customer relations finished stock control order servicing packing shipping creditors &amp; debtors cash flow payroll cost accounting profitability analysis personnel planning recruiting compensation policy</p>									

<그림 부.-1> CASA-SISP에 의한 정보체계 도출(가중치의 차=2, 그룹의 수=15)