최적 자원 및 운영관리를 지원하는 제조 데이터 분석시스템 라이브러리

| 저자 | Jayant Kalagnanam, Ph.D. Distinguished Research Staff Member / IBM Research
Young Min Lee, Ph.D. Research Staff Member / IBM Research
Tsuyoshi Ide, Ph.D. Senior Technical Staff Member / IBM Research
한형상 PD / KEIT 지식서비스 PD실

SUMMARY

★ 돌이킬 수 없는 대세로 자리잡은 IoT의 전방위적 보급 확산이 소기의 성과를 거두기 위해서는 두 가지 병목(스마트 센서, 데이터 분석 및 활용) 기술의 해결이 필수적임. 이 자료는 그 중 데이터 분석 및 활용기술에 관한 것으로, 이 부문의 선두주자인 IBM Watson 연구소가 개발하고 있는, 제조산업의 자원 및 운영관리와 관련된 다양한 문제들의 해결을 지원하는 데이터 분석시스템(Analytics) 라이브러리의 프레임워크 설명을 위한 것임. 빅데이터 기반의 이 분석시스템 라이브러리는 전자, 반도체, 자동차, 석유-화학, 오일-가스, 에너지-유틸리티, 그리고 광업 및 금속 등 다양한 제조산업이 주요 적용대상임

★ 제조산업용 분석시스템 라이브러리는 예지적 자산관리, 공정 및 장비의 예지적 고장관리, 공정 및 장비 모니터링, 데이터 분석 및 최적화, 예지적 환경분석시스템, 그리고 안전 및 효과적 운영을 위한 시공간적 분석과 같은 첨단 분석시스템들을 포함하고 있음. 분석시스템은 다양한 선진적 분석 알고리즘들과 포괄적인 온톨로지에 기초한 분류체계로 구성되어 있음. 용도별로 특화된 워크플로우, 공학적으로 설계된 API(Application Programming Interface)와 데이터 구조(Data Structure), 그리고 입력 데이터와 출력 솔루션의 가시화를 위한 GUI 도구들을 제공함으로써 일반 사용자들도 쉽게 사용할 수 있음. IoT3.0을 기반으로 센싱 데이터를 원격으로 수집할 수 있고, 비디오 · 텍스트 · 실시간 감지 데이터 등을 포함한 대규모의 다양한 구조적 · 非구조적 데이터(즉, 빅데이터)를 처리할 수 있도록 설계되었음

★ 이를 통해 이론적 분석 알고리즘들과 현장의 실제 문제들 사이의 간격을 메워주고, 정보기술(IT)과 운영기술(OT)의 융합을 촉진하며, 물리적 세계의 센싱 영역 확대를 통해 가속적으로 생성되는 데이터를 효과적 · 효율적으로 활용하여 Industry4.0의 비전 실현을 추구하고 있음

정책 시사점

★ 새로운 컴퓨팅의 시대에 데이터는 가장 중요한 천연자원으로 자리매김 되었음. 천연자원은 정련하여 용도에 맞게 잘 가공해야 가치를 발휘할 수 있음. 데이터가 기대대로 새로운 부가가치를 창출하고 성장동력의 일익을 담당할 수 있으려면 데이터의 수집 · 정비 · 분석이 적절히 이루어져 최종적으로는 사용자의 최적 의사결정을 지원할 수 있어야 함. 이를 위해 우선 요인분석(Factor Analysis)을 통해 데이터의 사용 목적에 맞춰 필요한 데이터의 종류와 수량을 미리 이해하는 게 중요함. 이를 통해 필요한 데이터는 반드시 확보하되, 데이터 수집을 위한 정보화 인프라 구축 시 과잉투자를 방지할 수 있음. 이같이 유용한 데이터의 확보기반 위에 영역별로 전문화된 분석시스템의 개발이 필수적임. 이러한 관점에서 여기에서는 제조산업의 자원 및 운영관리를 위한 분석시스템의 구조와 기능을 제시했음. 이 같은 접근방법은 제조산업뿐 아니라 마케팅 리서치, 에너지, 의료, 유통/물류, 국방, 레저, 문화, 교육, 연구개발 등 사회 모든 영역에 대해서도 유효한 것으로, 각기 특화된 전문적 분석시스템의 개발과 보급 지원이 필요함

1. 개요

★ 새로운 컴퓨팅의 시대가 도래하고 있음. 새로운 컴퓨팅의 시대란 인간처럼 감지하고 학습하며 습득된 지식을 기반으로 자동적으로 행동을 보정하는 방식으로 세상을 이해하는 인지적(Cognitive) 컴퓨터의 시대를 의미함. 1900년 무렵 나타난 1세대 컴퓨터는 산술계산용 기계들로 구성됨. 1950년대에 들어 프로그래밍이 가능한 2세대 컴퓨터가 출현했음. 인지적 컴퓨터는 많은 데이터, 기계학습, 그리고 단순 계산 이상의 정교한 분석시스템들로 복잡한 문제를 해결하는 전혀 새로운 접근방법을 제시함

★ 빅데이터는 그 동안의 성장도 엄청나지만 IoT, 소셜, 모바일, 그리고 클라우드와의 융합에 의해 계속 생성이 가속되고 있음. 산업체의 운영과정과 공정 및 설비들도 역시 막대한 규모의 데이터들을 생성하고 있음. 현재 1조에 달하는 지구상에서 연결된 사물과 기기들은 매일 2.5기가바이트의 데이터를 생성하고 있으며 그 속도는 가속되고 있음. 범세계적으로 기계들이 생성한 데이터의 양은 8.5ZB(제타바이트: 기가바이트의 1조배)에 달하며, 기하급수적으로 증가하여 5년 내(2020년까지) 40ZB에 이를 것으로 추정되고 있음[1]. 2020년까지 기계간 생성 데이터가 전체 데이터의 40%에 달하여 최대의 비중을 점유할 것으로 전망됨. 이미 센서와 기기들로부터 생성된 데이터의 양이 소셜미디어, 인터넷통화(VoIP), 그리고 기업 데이터보다 많아지고 있음

★ 데이터는 제조공장, 산업설비, 공장과 가정의 센서, 소셜 미디어, 스마트폰, 위성 이미지, 기상 연구실, 가전기기 그리고 심지어 우리가 운전하는 차량에 이르기까지 도처에서 다양한 사물로부터 생성되고 있음. 데이터의 유형도 구조적인 것, 非구조적인 것, 비디오, 오디오, 텍스트 등등 매우 다양함

★ 데이터는 새로운 자원으로 등장했고 산업은 데이터에 의해 혁명을 맞고 있음. 데이터를 적절히 활용함으로써 제조기업들은 품질을 개선하고 제조원가를 낮추며 공장의 안전도를 높이고 제조 에너지 효율을 개선하고 이익률을 개선할 수 있음. 이제 데이터는 제조 경쟁력의 원천으로 자리매김이 되었음. 데이터는 또한 스마트 건물, 스마트 도시, 스마트 지구의 구현을 도와 삶의 질 개선에도 기여함. 그러나 빅데이터의 수집 자체만으로는 아무런 도움이 되지 못함. 모든 데이터 원천들을 상호 연결하고 빅데이터를 수집하는 것은 중요하지만, 그 데이터를 처리하고 의사결정에 필요한 유용한 정보로 변환하는 적절한 분석도구가 없으면 빅데이터는 그저 큰 쓰레기에 불과함. 적절하게 분석되고 활용될 수 있어야 데이터가 산업과 사회에 도움이 될 수 있음. 새로운 컴퓨팅의 시대에 분석이 담당해야 할 역할은 기존의 분석과는 성격이 다름. 전통적인 분석은 더 이상 빅데이터의 양과 종류, 생성속도, 진실성(검증) 요구를 감당할 수 없음. 새로운 컴퓨팅의 시대에 필요한 것은 시스템의 구축과, 사양 설정 및 운용이 용이하고, IoT 및 빅데이터와 총체적인 통합을 손쉽게 이룰 수 있는 선진적인 분석시스템임. 이러한 선진적 분석시스템은 고급 수학, 통계학, 데이터 마이닝, 기계학습, 인지적 컴퓨팅, 그리고 최적화 같은 다양한 알고리즘과 기술의 복합적 집합체로 구성됨

★ 새로운 컴퓨팅 시대에 산업부문이 선진적 분석시스템의 혜택을 누릴 수 있는 중요 기술 영역의 하나가 자원 및 운영관리로, 이는 제조산업의 자산, 자원, 운영 및 공정과 관련된 핵심 문제들의 집합임. IoT와 통합하고 빅데이터를 저장, 처리, 분석하는 능력이 다양한 자원 및 운영관리 문제들의 개선과 산업 경쟁력 향상에 도움을 줄 것임

★ 빅데이터와 IoT를 분석시스템과 함께 활용함으로써 다양한 산업의 공정운영과 관리를 효과적, 효율적으로 개선할 수 있음. 그러나 그러한 분석시스템의 개발에는 상당한 어려움이 따르는 바, 실용성 있는 솔루션의 개발을 위해서는 대상 분야에 대한 경험과 함께 데이터 마이닝, 기계학습, 통계분석 및 최적화 이론 전반에 대한 전문적 지식이 필요하기

때문임. 따라서 분석시스템의 개발에는 상당한 자원과 시간의 투자가 요구되므로 이를 직접 개발하여 활용할 수 있는 기업은 별로 없음. 수리적 모형이나 통계분석에 특화된 교육을 받지 않은 일반 사용자가 특정 문제의 해결에 적합한 분석시스템을 기획하고, 적합한 과정과 방법으로 개발하며, 이를 활용하여 소기의 성과를 낼 것을 기대하기는 어려움. 데이터 자체의 복잡성 증가와 더불어 적합한 분석 알고리즘을 선별하는 일도 갈수록 IoT 시대의 또 다른 도전적 과제로 대두하고 있음

★ 이러한 관점에서 SROM(Smarter Resource and Operations Management) 이라는 이름으로, 다중적 계층구조로 설계된, 선진적인 자원 및 운영관리를 위한 분석시스템 라이브러리를 제공하는 플랫폼을 개발하고 있음. 이 플랫폼은 특화된 수리적 모형 기법의 교육을 받은 적 없는 일반 사용자들도 적은 시간과 노력만으로 손쉽게 산업계의 복잡한 자원 및 운영관리 문제의 해결에 적합한 분석시스템의 선별, 개발, 활용을 가능하게 해 줄 것임

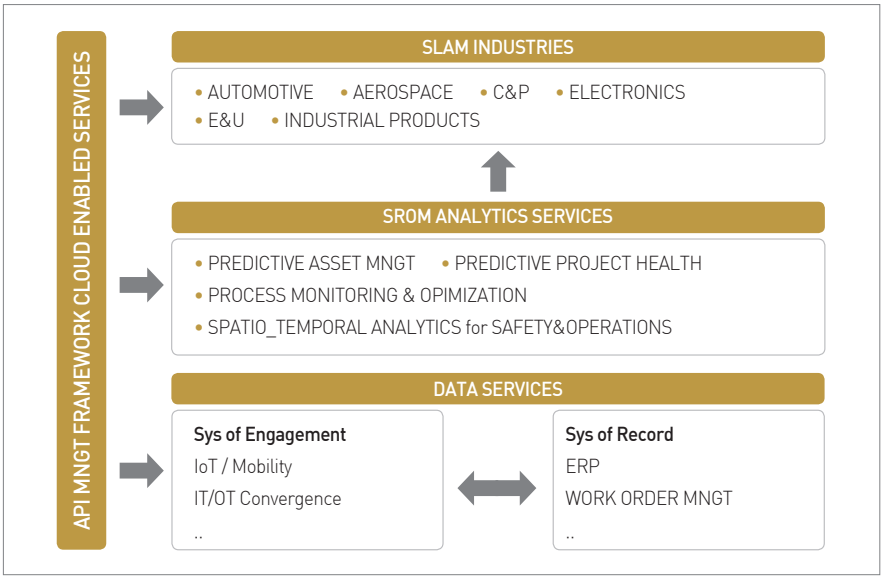
2. SROM의 성능

★ SROM은 자원 및 운영관리 분야에서 현실과 분석이론 간의 간격을 메워주는 차세대 분석시스템의 프레임워크임. 산업계의 문제 해결을 위한 대부분의 기존 분석시스템들은 실제 현장적용 사례와 산업 특성에 맞는 솔루션이 부족함. 산업계의 실제 문제와 이론적 알고리즘 간에는 상당한 거리가 있음. SROM은 풍부한 선진적 알고리즘들로 뒷받침되는 분석시스템들의 라이브러리를 종합적으로 제공하고 있음. 기업 시스템용 분석 플랫폼들은 순수한 분석기능 엔진에서 산업별 목적에 맞춰 특화된, 사업/기업 기능지향적 솔루션으로 진화하고 있음. 1세대 분석시스템들은 전형적으로 순수한 통계분석 기능과 기계학습 라이브러리로 구성되었는 바, 예로 S, R, Weka 등이 있음. 2세대 분석시스템들은 (사용자가) 구성할 수 있는 GUI 시스템을 구비한 기계학습 라이브러리로 구성되었고, 예로는 SAS, SPSS 등이 있음. 3세대 분석시스템은 클라우드를 기반으로 하고, 솔루션 지향적이며, 기능기반의 API를 제공하고 있는 바, SROM도 그 중의 하나임. SROM은 자산집약적 산업에서 빠르게 변화하고 있는 솔루션 수요에 대응할 수 있도록 설계됨

★ SROM은 온톨로지 기반의 분석시스템 라이브러리로서, 빅데이터 및 IoT와 통합운영이 가능하도록 설계되었음. SROM은 여러 산업분야에서 활용될 수 있게 설계되어 전자, 반도체, 자동차, 석유-화학, 오일 및 가스, 에너지 및 유틸리티, 광업 및 금속산업과 같은 제조 산업에서 활용이 가능함. 분석시스템들은 다양한 선진적 알고리즘들의 뒷받침을 받고 있고, 포괄적인 온톨로지에 따라 분류되며, 공학적으로 설계된 API, 입력 데이터와 출력 솔루션의 가시화를 위한 데이터 구조 및 GUI 도구, 그리고, 용도별로 특화된 워크플로우를 제공함으로써 일반 사용자들의 활용을 지원하고 있음

★ 분석 라이브러리는 Figure 1에 표시된 바와 같이 예지적 자산관리, 공정 모니터링 및 최적화, 그리고 효과적인 안전관리 및 운영을 위한 시공간적 분석 같은 선진적 분석시스템들을 포함하고 있음. SROM은 클라우드 환경에서 제공되는 서비스로서 데이터 서비스와 통합되어 있는 바, 이와 관련된 데이터의 유형은 두 가지가 있음. 첫째 유형은 산업체 운영과정에서 센서와 장비로부터 동적으로 수집되는 IoT 데이터임. 이러한 데이터는 IT 시스템뿐 아니라 SCADA (Supervisory Control And Data Acquisition) 시스템, DCS(Distributed Control System), MES (Manufacturing Execution System)와 기타 OT(Operations Technology) 시스템들로부터 생성됨. 둘째 유형의 데이터는

ERP(Enterprise Resources Planning), 작업순서관리, 자산관리 등과 같은 시스템들의 기록 데이터임. SROM 분석시스템은 이 두 가지 유형의 데이터를 모두 처리할 수 있음. 대상 문제에 따라서는 데이터 크기는 매우 커질 수 있고(즉 빅데이터), 비디오, 텍스트, 실시간 센서 데이터 같이 구조화 혹은 구조화 되지 않은 데이터일 수도 있음. Figure 1의 중앙에 위치한 SROM 분석 서비스 계층은 산업별로 패키지화가 가능함 - 예컨대 자동차 산업용, 석유-화학 산업용, 에너지 및 유틸리티 산업용 분석 라이브러리 패키지 등



| Figure 1 SROM의 개관 |

★ 자원 및 운영관리 분야에서 다양한 산업계 문제들의 해결에 적용이 가능한 분석시스템들의 유형은 다양하지만, 일반 사용자들이 특정 문제의 해결에 적합한 분석시스템과 알고리즘을 식별해 내는 것도 어렵지만, 시범 적용을 해 보거나 동종 산업 내 다른 회사에 적용된 사례를 이용하는 것도 쉽지 않음. SROM의 분석시스템들은 분석시스템에 특화된 온톨로지에 기반하여 분류되어 있음. 이 온톨로지는 수리적 모형 기법의 교육을 받은 적 없는 일반 사용자들도 산업별로 유용한 분석시스템과 분석시스템간의 연관성을 이해하고, 산업체의 복잡한 자원 및 운영관리 문제의 해결에 적합한 분석시스템을 손쉽게 식별하고 궁극적으로는 개발도 가능하게 함. Figure 2는 SROM 온톨로지의 전체 구조를 단순화 해서 보여주고 있음

★ Figure 2에 표현된 것과 같이 자원 및 운영관리 분야의 SROM 온톨로지에는 4가지 수준의 요소들이 있음. 최상위 수준(레벨 1)은 분석시스템들의 그룹으로서 이에는 예지적 설비보전, 예지적 고장분석, 공정 및 장비분석, 그리고 공정 모니터링 및 최적화가 있음

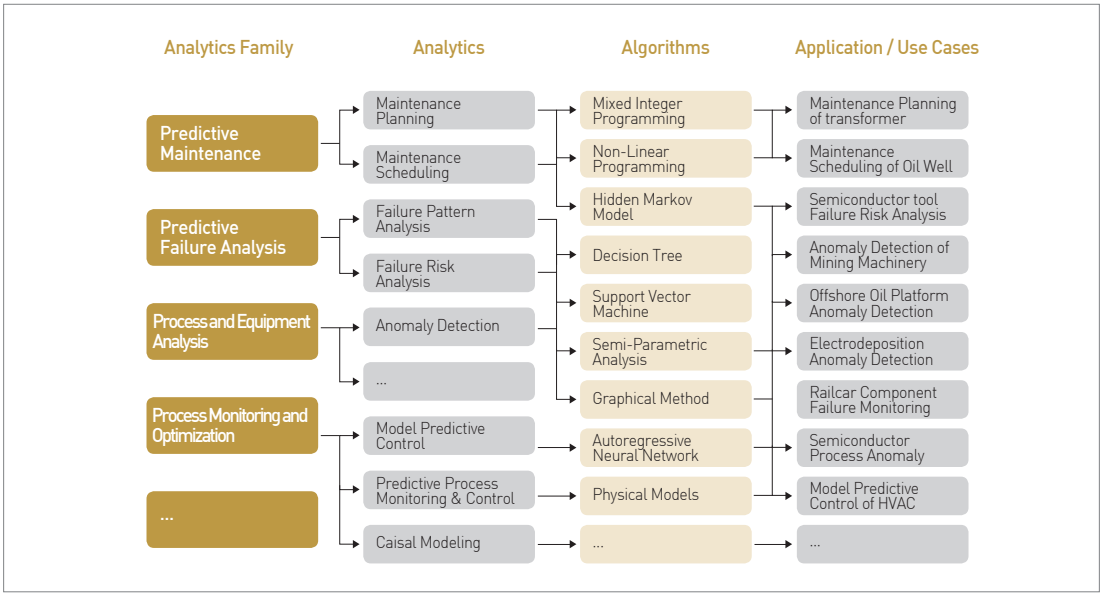
★ **예지적 설비보전 분석시스템(Predictive Maintenance Analytics)**은 다양한 자산 및 장비의 공정 및 운영 조건을 결정하고 설비보전이 필요한 시점을 예측해 줌. 설비보전이란 장비를 검사하고, 장비나 부품의 수리 및 교체를 수행하는 작업임. 이러한 분석시스템은 보통 혼합정수-선형계획법(MILP: Mixed Integer Linear Programming) 으로 모형화 되는 최적화 문제로, 목적함수는 유지보수와 고장에 따른 비용의 최소화이며, 제약조건은 가용 자원 · 부품 · 예산 그리고 운영임. 예지적 설비보전 분석시스템은 유지보수 비용의 감축, 장비의 사용수명 연장, 제품의 품질 향상, 그리고 공장의 안전도 향상에 도움이 됨. 예지적 설비보전 분석시스템에서, 예지보전계획수립(Predictive Maintenance Planning)은 장기계획 수립을 위한 것으로, 대개 자산의 고장에 따른 위험 요인을 고려하여 1년 이상의 기간에 대해 주단위 혹은 월단위로 설비보전계획을 수립함. 보전일정계획수립(Maintenance Scheduling)은 설비보전을 위한 단기 일정계획 수립을 위한 것으로, 대개 하루 혹은 1주정도의 기간에 대해 시간 혹은 분단위로 보전 작업자, 자원, 경로를 포함한 작업일정 계획을 수립함

★ **예지적 고장분석(Predictive Failure Analysis)**은 자산이나 장비의 미래 고장발생을 예측하여 고장 방지를 위한 예방조치를 취하기 위한 것임. 생산공정에서 자산이나 장비의 고장은 제품 품질, 생산원가, 안전도 등에 중대한 부정적 영향을 미칠 수 있음. 예지적 고장분석에는 두 가지 유형이 있음. 고장유형분석(Failure Pattern Analysis)은 고장의 전조가 되는 조건(유형)을 찾아내는 것이고, 고장위험도분석(Failure Risk Analysis)은 고장 이력 및 운영조건 데이터를 분석해서 자산의 예상 수명을 추정하는 것임

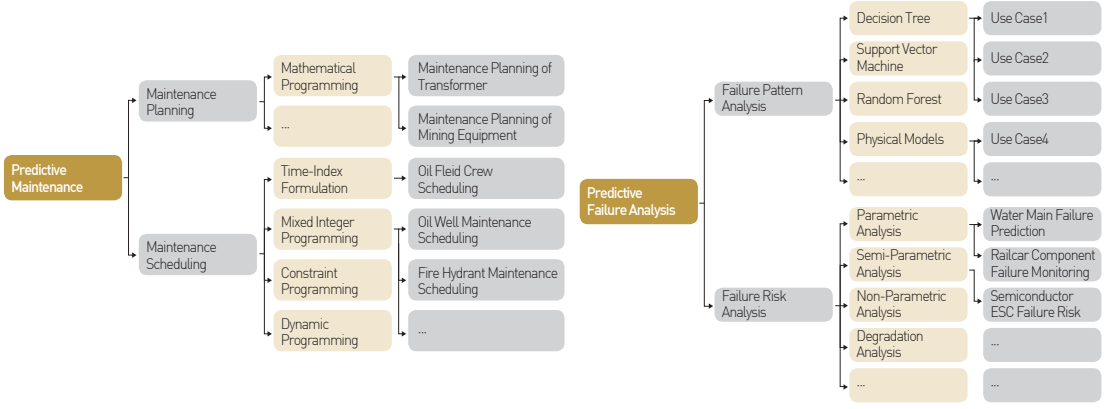
★ **공정 및 장비분석(Process and Equipment Analysis)**은 공정 및 장비로부터 생성되는 센싱 데이터들의 흐름을 분석해서 이상거동 상황이나 오류를 찾아내기 위한 것임. 그래서 이를 고장 감지 및 진단(FDD: Fault Detection and Diagnosis)이라고 부르기도 함. 좋은 예로 정상 상태라고 규정된 한계를 벗어나는 장비의 이상 거동을 찾아내는 이상거동 검출(Anomaly Detection)이 있음. 이상거동 검출은 우선 다양한 조건에서 운영된 장비의 과거 거동 데이터를 분석하여 장비의 정상적인 거동 범위를 정의하고, 이러한 정상범위를 벗어나는 거동을 찾아내어 잠재적인 이상거동으로 규정한 뒤 경보를 생성함

★ **공정 감시 및 최적화(Process Monitoring and Optimization)**는 지속적으로 공장 및 장비를 감시하고, 공정이나 작업의 산출을 극대화 하면서 가장 비용 및 에너지 효율이 높고 안전한 운영이 되도록 제어변수의 값을 최적화 해 주는 일련의 분석시스템들로 구성됨. 이를 위하여 우선 공정이나 작업의 예측모형(물리적 모형, 데이터기반 모형, 혹은 혼합 모형이든 관계없이)의 개발이 필요함. 다음에는 이러한 예측모형이 산출한 예측치를 이용하여 결정변수들의 최적 수치들을 계산해 주는 최적화 모형 개발이 필요함

★ 각 분석시스템 그룹에는 다수의 분석시스템들(레벨 2)이 포함되어 있음. 예컨대, 예지적 설비보전 그룹에는 보전계획수립과 보전일정계획분석 등이 포함되어 있음. 예지적 고장분석에는 고장유형분석과 고장위험도분석이 포함됨. 분석은 다양한 알고리즘들(레벨 3)에 의해 실행됨. 예컨대, 고장유형분석은 3가지 유형의 알고리즘, 즉 Decision Tree, Support Vector Machine, 그리고 Random Forest에 의해 실행됨. 마지막 수준(레벨 4)에는 적용사례의 샘플 어플리케이션들이 있음. 예컨대, 이상거동 검출용 그래픽 방법론(Graphic Method of Anomaly Detection)을 위하여 분석시스템은 광산기계용 및 해양 오일 플랫폼용 이상거동 검출 적용사례를 제공함. Figure 3는 예지적 설비보전 및 예지적 고장분석시스템들의 보다 심층적인 계층구조를 보여주고 있음. 여기에 제시된 온톨로지는 개발 과정에 있는 것으로 아직 완성된 것은 아님. 온톨로지는 더 많은 산업계의 문제를 처리해 가는 과정에서 지속적으로 간소화 되고 개선될 것임. 제 6장에서 일부 분석시스템, 알고리즘, 그리고 적용사례들이 제시될 것임



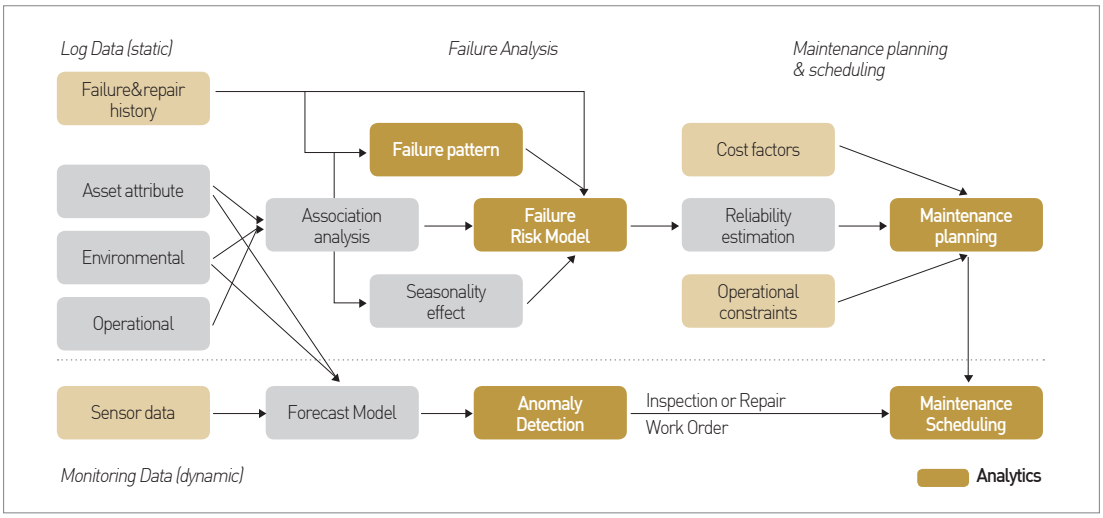
| Figure 2 SROM 온톨로지의 개관 |



| Figure 3 예지적 설비보전 및 예지적 고장분석의 온톨로지 |

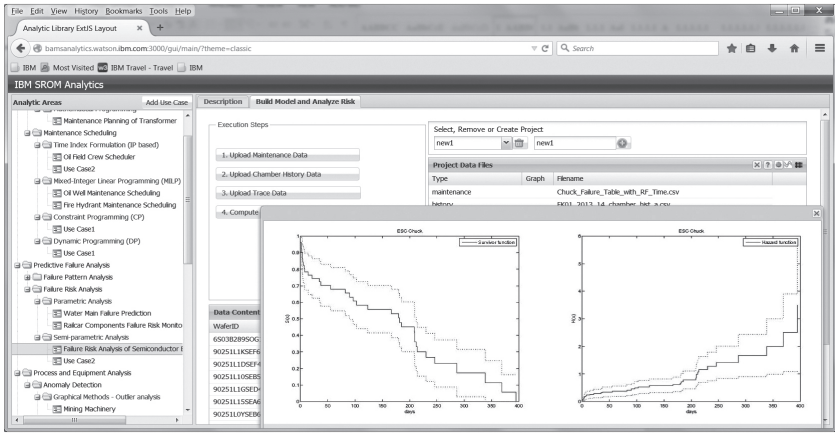
★ 각 분석시스템은 미리 정해진 입력 데이터를 사용하여 솔루션으로 특정 출력 데이터를 생성함. 더하여 한 분석시스템의 결과는 다른 분석시스템의 입력 데이터로 이용될 수 있음. Figure 4는 분석시스템과 관련 데이터간 관계의 한 예를 보여주고 있음. 이 사례에서 고장 이력 데이터, 자산의 속성, 그리고 환경 및 운영 데이터가 고장유형분석 모형과 고장위험도분석 모형에 이용되고 있음. 고장유형분석 모형의 출력물, 설비보전 및 고장의 비용요소와 위험 함수(Hazard

Function), 신뢰도 추정과 설비보전 예산 및 자원 같은 작업 제약조건들이 유지보수계획 최적화 모형의 입력데이터로 사용됨. 유지보수계획 및 검출된 이상거동 사례는 유지보수 일정계획 모형의 입력 데이터로 사용될 수 있음



| Figure 4 분석시스 템과 데이터간 관계 |

★ 분석시스템의 온톨로지는 Figure 5의 좌측에 표현된 것처럼 웹기반 GUI 에서 쉽게 운용될 수 있도록 계층적 메뉴로 전개되어 있음



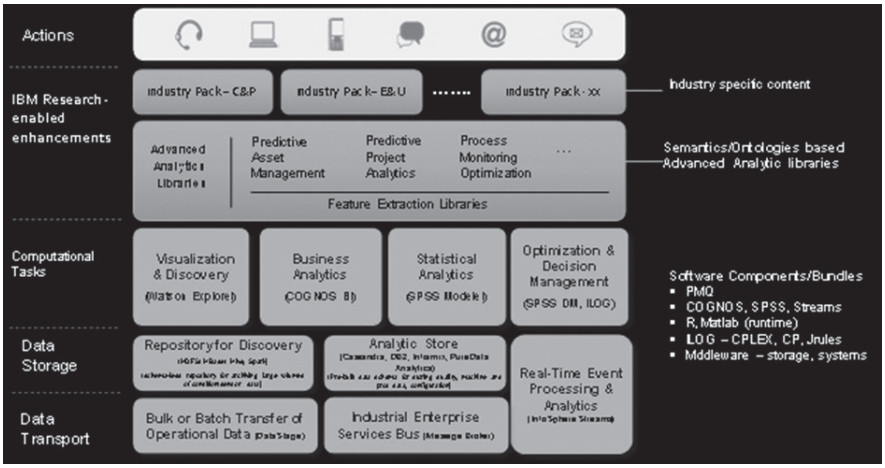
| Figure 5 시맨틱 기반으로 정의된 SROM 분석시스템 솔루션의 사례 |

3. SROM의 아키텍처

★ SROM의 핵심은 Figure 6에서 표현된 것처럼 예지적 자산관리, 공정 모니터링/최적화 등의 분석시스템 그룹으로 구성된 선진적 분석시스템 라이브러리임. 이 라이브러리는 용도별로 일부분을 분리해서 특정 산업을 위한 산업용 패키지로도 만들 수 있음. 예컨대, Industry pack-C&P는 SROM의 석유-화학산업용 버전이고, Industry pack-E&U는 에너지 및 유틸리티 산업용 버전임. 분석시스템은 시맨틱 및 온톨로지 기반으로 설계되었기 때문에 일반 사용자들도 온톨로지의 계층을 보고 자신이 가진 특정 문제의 해결에 적합한 분석시스템을 손쉽게 찾을 수 있음

★ SROM 분석시스템들은 Computing 엔진들을 보유한 계산(Computational) 계층의 지원을 받고 있음. 최적화 및 의사결정관리가 포함된 분석시스템을 위하여 최적화 엔진인 ILOGTM Cplex, 최적화 플랫폼인 ILOG Decision Optimization Center (DOC), 그리고 SPSSTM DM 같은 의사결정관리 도구가 사용될 수 있음. 통계적 분석을 포함한 분석시스템을 위해서는 SPSS Modeler, R, Python 이나 Matlab 같은 통계 엔진을 사용할 수 있음. 비즈니스 분석을 위한 분석시스템을 위해서는 COGNOSTM BI 같은 사업데이터 분석도구를 이용할 수 있음. 가시화 및 발견을 위한 분석시스템 용으로는 예컨대 Watson Explorer 가 사용될 수 있음

★ 컴퓨팅 엔진은 비디오, 텍스트, 실시간 센싱 데이터 같은 다양하고 규모가 큰 구조적, 非구조적 데이터(빅데이터)를 처리하고, 이는 데이터 저장(Storage) 계층의 지원을 받고 있음. 데이터 저장 계층은 대량의 운영 및 센싱 데이터의 저장을 위하여 HDFS(Hadoop Distributed File System), HBase, Hive, Spark 같은 데이터 리파지토리(저장소)를 가지고 있음. 또한 품질, 기계 및 제품 데이터, 사양의 저장을 위한 데이터 스키마와 함께 Cassandra, DB2, Informix, PureData Analytics 같은 데이터베이스 시스템을 가지고 있음. IoT 응용을 위해서는 InfoSphere Streams 같은 실시간 이벤트 처리 기능도 활용이 가능함



| Figure 6 시맨틱 기반으로 정의된 SROM 분석시스템 솔루션의 아키텍처 |

★ 데이터 저장 계층은 다양한 데이터 원천으로부터 운영 데이터의 대량 혹은 일괄 전송을 실현하는 데이터 전달(Data Transport) 계층의 지원을 받고 있음

★ SROM은 클라우드 기반의 솔루션이며, SROM 분석시스템에 의해 계산된 결과는 실행(Action) 계층에서 컴퓨터 모니터, 모바일 기기, 이메일, 소셜 미디어, 음성 등으로 사용자에게 전달될 수 있음

4. SROM 과 IOT 및 빅데이터

★ 사물인터넷(IoT)은 미래의 컴퓨터 및 통신 분야에서 시간과 장소에 관계없이 모든 사물을 연결한다는 개념에 기초한 기술 혁명임[5]. IoT는 인터넷으로 접속되는 사물 및 기기들의 네트워을 나타내는 것으로, 그러한 사물들로는 공장 장비, 가전제품, 센서, 측정장비, 스마트폰 같은 것들이 있음. 우리는 이미 일상생활에서 IoT 솔루션들을 실제 사용하고 있음. 예컨대, 스마트폰을 통하여 가정 보안장치, 조명, 난방, 에어컨, 기타 가전제품들의 상태를 파악할 수 있음. 제조환경에서는 IT 시스템뿐 아니라 다양한 SCADA(Supervisory Control And Data Acquisition) 시스템, DCS(Distributed Control System), MES (Manufacturing Execution System), 그리고 기타 운영 시스템들로부터 더 많은 데이터가 더 빠른 속도로 생성되고 있음. IoT는 자동차, 석유-화학, 광업, 금속, 전자 등 거의 모든 산업에 영향을 미치고 있음

★ 빅데이터는 다양한 원천으로부터 생성되는 데이터의 급속한 축적이라는 성격을 가진 하나의 현상임. 데이터가 워낙 급속히 생성되고 축적되는 까닭에 산업과 사회에 데이터가 범람하고 있음. 데이터는 산업에게 가치있는 새로운 천연자원임. 우리는 데이터를 새로운 상품과 서비스의 창출이나 기존의 상품과 서비스의 개선에 활용할 수 있음. 따라서 빅데이터는 도전이자 동시에 기회이기도 함. 도전은 이 빅데이터를 어떻게 잘 길들일 수 있는가의 문제이고, 기회는 어떻게 이 정보를 적절히 분석하고 활용해서 산업과 사회에 효과적으로 기여할 것인가의 문제임. 현재 빅데이터 솔루션들은 다음 4가지 관점에서 전통적인 IT 솔루션들과 구별해서 보는 것이 일반적임

- 부피 : 빅데이터 솔루션들은 대량의 데이터를 관리 및 처리할 수 있어야 함
- 속도 : 빅데이터 솔루션들은 더 빠르게 전달되는 데이터를 처리할 수 있어야 함
- 종류 : 빅데이터 솔루션들은 데이터의 구조성 여부와 관계없이 더 많은 종류의 데이터를 처리할 수 있어야 함
- 진실 : 빅데이터 솔루션들은 급속히 전달되는 대량의 데이터를 검증할 수 있어야 함

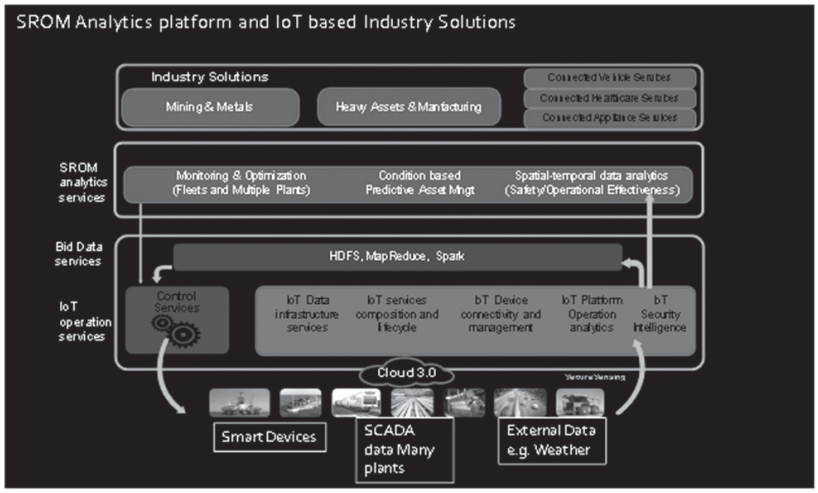
★ 빅데이터는 우리 주변의 모든 것들로부터 상시적으로 생성되고 있음. 모든 디지털 프로세스와 소셜 미디어가 데이터를 생성하고 있고 시스템, 센서, 모바일 기기가 데이터를 전달하고 있음. 빅데이터는 다양한 원천들로부터 놀라운 속도, 양, 다양성, 진실성으로 전달되고 있음

★ 빅데이터와 IoT는 우리가 환경과 함께 일하고 놀고 상호작용 하는 방식의 성격에 근본적인 변화를 가져오는 컴퓨팅 패러다임임. 빅데이터가 부피, 속도, 종류, 진실성에 대한 문제라면 IoT는 생산성과 삶의 질 개선을 위하여 그 데이터를 의미있게 활용하는 방법에 대한 문제임[6]. 그러나 빅데이터나 IoT 자체만으로는 쓸모가 없음. 데이터가 적절하게 분석되고 유용한 정보로 변환될 경우에만 빅데이터의 장점이 실현될 수 있음. 분석시스템은 빅데이터로부터

유용한 가치를 이끌어내기 위한 열쇠임. 빅데이터는 조직내 구성원들이 함께 일하는 방식도 변화시키고 있음. 모든 데이터로부터의 가치 실현을 위해 사업부문과 IT 리더들이 힘을 하나로 모으게 하는 문화를 만들고 있음. 빅데이터로부터 얻어지는 통찰력은 모든 구성원들로 하여금 고객참여도 심화, 제품품질 개선, 운영 최적화, 위험과 부정의 예방, 그리고 새로운 수입원의 현금화 같은 문제들에 대해 보다 나은 의사결정을 내릴 수 있게 함[2]. 그러나 이 모든 것들은 분석시스템의 효과적인 활용을 통해서만 가능한 것임. 분석시스템은 선진적인 수학, 통계학, 데이터 마이닝, 기계학습, 인지적 컴퓨팅의 집합체임. 범-산업적 분석시스템 라이브러리, 즉 SROM의 목적은 산업계로 분석 능력을 신속히 보급하는 데 있음

★ SROM은 자원 및 운영관리 분야의 다양한 문제들을 해결해 주는 분석시스템들을 내장한 라이브러리이자 개발용 플랫폼임. 플랫폼은 일반 사용자들이 수리적 모형에 관한 교육을 받은 적이 없더라도 자원 및 운영관리 분야의 복잡한 산업계 문제해결을 위한 분석시스템을 손쉽게 선별, 개발, 활용할 수 있게 해 줌

★ Figure 7은 IoT, 빅데이터와 함께 SROM 플랫폼을 설명하고 있음. SROM을 위한 데이터의 원천은 Figure 7의 하단에 제시된 것과 같이 스마트폰이나 태블릿PC 같은 스마트 기기, SCADA 시스템 같은 산업체 공장의 IT 및 OT 시스템, 그리고 날씨 데이터, 지리 데이터, 교통 데이터 같은 외부 원천들이 있음. 이러한 IoT 데이터들은 IoT 운영 서비스 계층에서 지원되며, 여기에는 IoT 데이터 인프라 서비스(IoT Data Infrastructure Services), IoT 서비스 구성 및 라이프사이클 서비스(IoT Services Composition and Lifecycle Services), IoT 기기 연결/관리 서비스(IoT Device Connectivity/ Management Services), IoT 플랫폼 운영 분석시스템 서비스(IoT Platform Operation Analytics Services), 그리고 IoT 보안정보 서비스(IoT Security Intelligence Services) 등이 포함되어 있음. IoT 운영서비스 계층에서 수집된 데이터는 빅데이터 서비스 계층에서 모아지고 분배되며, 이에는 HDFS, Spark, MapReduce 등이 있음. 빅데이터 서비스 계층은 다시 SROM 분석시스템 계층에서 접속되며, 이에는 자원 및 운영관리 분야의 다양한 분석시스템들의 집합체가 포함되어 있음. SROM 분석시스템은 광산 및 금속, 석유-화학, 전자 등 여러 산업군별 솔루션 패키지로 분리할 수 있음



| Figure 7 SROM 분석시스템 플랫폼과 IoT/빅데이터 기반의 산업용 솔루션 |

5. SROM의 배치

★ SROM 분석시스템들의 운영을 위해서는 데이터 인터페이스 프로그램, 데이터베이스, 데이터 통합 및 변환, 데이터 웨어하우스, 통계분석 엔진, 최적화 엔진, 그리고 가시화 등 매우 정형화된 컴퓨팅 HW 및 SW를 필요로 함. 또한 처리해야 할 데이터의 크기와 분석의 양에 따라서 이들 컴퓨팅 자원의 증감이 필요함. 일반 사용자들은 개인적으로 그러한 자원을 보유할 수 없음. 그래서 클라우드는 많은 사용자들이 모든 자원들을 비용대비 효과가 높고 안전한 공유를 통해 이러한 분석 서비스를 안전하고, 유연하며, 신축성 있게 제공하는 대안이 되고 있음. 공유된 컴퓨팅 자원을 이용하여 클라우드 상에서 다수의 사용자들에게 이러한 분석시스템 서비스를 제공하는 것은 전통적인 컴퓨팅 수단을 통하여 개별 사용자에게 같은 서비스를 제공하는 것보다 훨씬 비용대비 효율이 우수함. 사용자의 숫자와 목적에 따라 적합한 컴퓨팅 자원의 수준을 클라우드 상에서 쉽게 가상으로 구성할 수 있고, 자원의 활용수준도 수요에 따라 보정이 가능함. 클라우드를 이용하면 사용자들은 어떤 특별한 HW, SW 인터페이스, 수리적 도구를 설치하지 않아도 항상 모든 분석시스템들을 웹 브라우저를 통하여 활용할 수 있음

★ SROM은 분석시스템 서비스(AaaS: Analytics as a Service) 모형으로 설계가 되었음. SROM은 서비스의 주체이며 인터넷을 통해 사용자들에게 서비스됨. 클라우드 상에 SROM을 배치함으로써 사용자의 컴퓨터에 분석시스템을 설치하고 가동시켜야 하는 필요성을 없앴고, 더불어 SW 유지보수, 지속적인 가동 및 지원을 받아야 하는 부담을 경감시켜 줌. SROM을 클라우드에 배치함으로써 동적인 작업부하의 처리를 위하여 강력한 보안이 제공되는 기업 레벨의 가상 서버 환경을 신속히 구축 · 활용할 수 있음. SROM 클라우드는 접근을 제어하고 보안을 설정할 수 있는 능력을 제공함

6. SROM의 분석시스템

SROM은 아직 개발 중이며 계속 분석시스템과 적용사례가 추가되고 있어 일부 분석시스템에 대해서만 선별적으로 설명을 하겠음

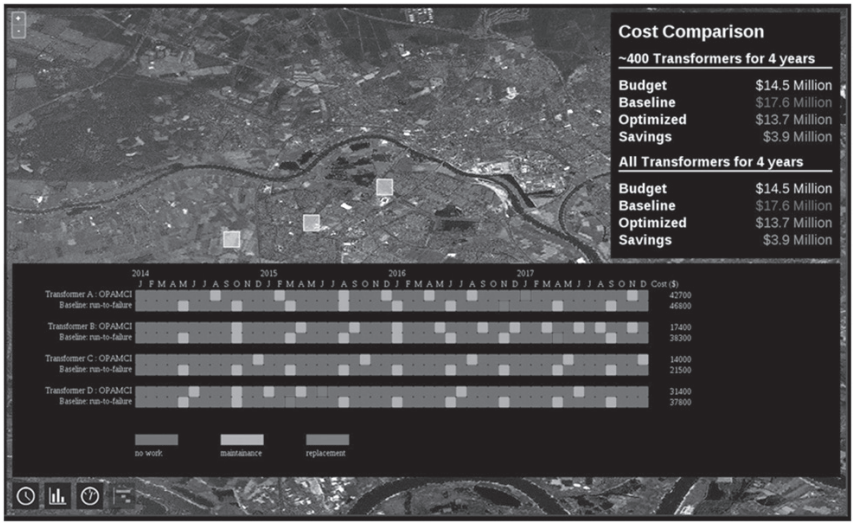
6-1. 유지보수계획 수립(Maintenance Planning)

유지보수계획 수립은 한 자산의 전체 생애주기를 고려하여 비용이 최소화되는 유지보수 계획의 수립을 목적으로 함. 이 분석은 수리, 고장시간, 교체의 비용을 고려한 최적의 수리 및 교체 계획을 생성함. 계획수립의 대상 기간은 그 자산의 전체 생애주기이며, 따라서 값비싼 자산의 경우에는 수 십 년이 될 수도 있음. 유지보수계획 수립은 최적 유지보수, 수리, 교체 시기를 정하는 최적화 실행문제임. 계획수립 실행을 위한 입력 데이터로 사용 이력과 조건정보를 기반으로 자산의 예상수명을 예측하는 고장 위험도 분석(즉, 생존함수: survival function)을 필요로 함

적용사례 : 변압기의 유지보수계획 수립

★ 이 적용사례는 가용 예산 및 인력 같은 제약조건을 고려하여 특정 집합의 자산들을 대상으로 최적의 예방보전/교체 일정을 계산해 낸 사례임. 이 모형의 입력 데이터는 자산정보, 고장률 함수(Hazard Function) 데이터, 그리고

유지보수를 위한 가용 예산 및 인력 같은 제약조건임. 이 적용사례는 한 유틸리티 회사를 위한 변압기의 최적 유지보수 계획을 수립한 것임. 이 최적화 문제는 非선형계획(NLP: Non-Linear Programming) 및 혼합정수계획(MILP: Mixed Integer Programming) 모형으로 정형화 되어 해결되었음



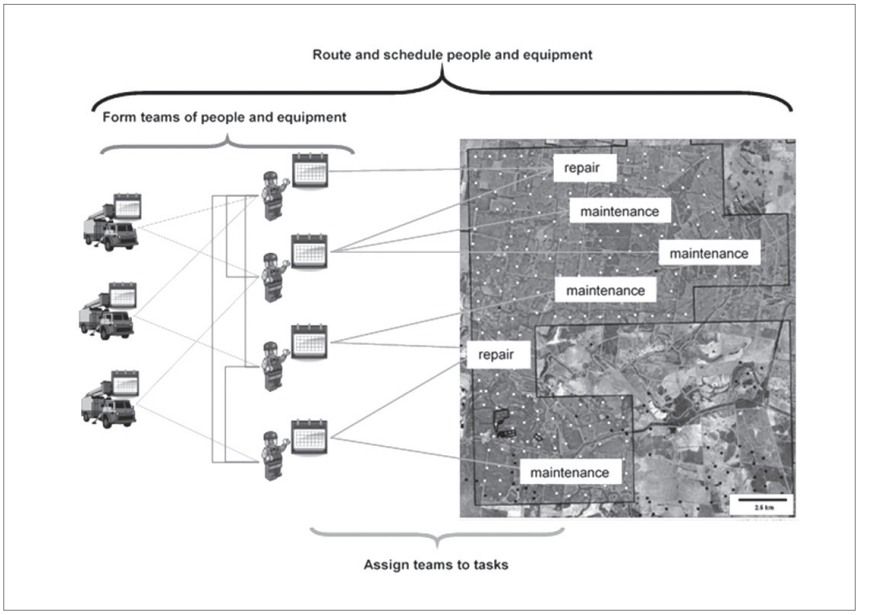
| Figure 8 유지보수계획 최적화 분석시스템의 한 출력 결과물 |

6-2. 유지보수 일정계획 수립(Maintenance Scheduling)

유지보수 일정계획 수립은 주 내지 월 단위의 기간을 대상으로 하는 운영계획 수립문제임. 다양한 자산을 보유한 공장을 대상으로 하며, 개별 자산은 각기 생애주기에 따른 유지보수 계획이 있음. 이제 특정한 달(혹은 주)에 자산별로 미리 정해진 검사/수리 작업 지시가 있는지 여부를 확인할 수 있음. 주어진 기간에 수행해야 하는 모든 활동들을 집계한 뒤, 작업자 · 차량 · 부품재고 · 공구 등의 제약조건을 감안해서 작업순서를 정하는 일정계획 수립문제를 풀어야 함. 이러한 일정계획 수립 시에는 자산의 위치와 이동 거리도 고려 대상에 넣어야 함. 일정계획 수립 실행을 위해서는 작업순서, 필요자원, 작업실행이 필요한 시간대 등의 입력 데이터가 필요하고 유지보수를 위한 비용 최소화 일정계획을 생성해 주어야 함

적용사례: 유전의 유지보수 일정계획 수립

★ 이는 유전의 유지보수 및 검사를 위한 최적의 유지보수 일정계획과 자원(즉, 인력, 장비, 물자 등)의 경로계획(Routing) 수립을 통해 최소 비용으로 유지보수 수요를 충족시킨 사례임. 시스템은 지리적으로 산재되어 있는 임무 수행을 위하여 인력과 자원들의 경로와 작업일정 계획을 생성함. 사용자는 임무의 위치와 소요시간을 입력해야 하는데, 소요시간은 임무수행에 투입된 장비와 기술자의 유형에 따라 달라질 수 있음. 시스템은 혼합정수계획 모형을 이용하여 이동 거리와 총 시간을 최소화 하면서 주어진 기간 동안 최대로 많은 임무 수행이 가능한 솔루션을 찾아내게 됨



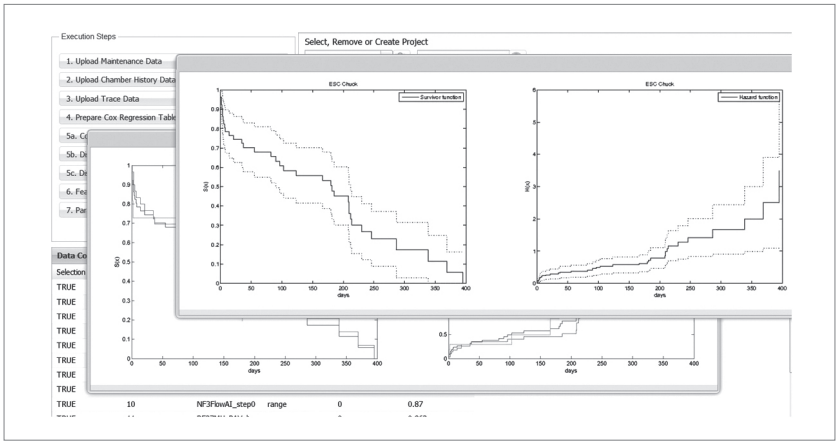
| Figure 9 유지보수 최적화 분석시스템의 개관 |

6-3. 고장위험도분석(Failure Risk Analysis)

고장위험도분석은 같은 등급의 자산에 대한 고장이력 데이터를 기반으로 자산의 예상수명을 추정함. 한 자산의 예상수명은 자산의 이력과 속성, 그리고 환경 및 운영 스트레스 요인에 의해 결정됨. 일반적으로 고장위험도 모형은 자산의 생존율을 다양한 환경 및 운영요소와 자산 속성들의 함수로 나타냄. 생존율은 자산의 이력을 기반으로 자산이 미래 일정 기간 동안 고장없이 가동될 확률을 계산함. 생존분석은 파라메트릭, 부분-파라메트릭, 그리고 非-파라메트릭 기법들로 구성됨. 생존연구에서 가장 일반적으로 사용되고 있는 생존함수나 고장률 함수 같은 척도의 값을 추정하고, 그룹별로 이들을 비교하며, 생존시간에 대한 예측변수들의 영향을 분석함

적용사례 : 반도체 정전기 척(ESC: Electrostatic Chuck)의 고장위험도분석

★ 이 적용사례는 반도체 제조공정에서 정전기 척의 예상수명을 예측한 것임. 유지보수 이력 데이터, 제조장비의 운영 이력 및 센싱 데이터 등을 분석해서 이 시스템은 생존/고장률 함수를 예측하고, 어떤 공정 변수들이 척의 고장에 어떤 영향을 주는지를 분석하고, ESC의 유지보수와 교체가 필요한 조건을 찾아냄. 그러면 위험도분석은 유지보수 및 고장 비용을 최소화 해 주는 최적의 유지보수 및 교체계획을 생성함. 이 분석은 콕스 비례위험회귀(Cox Proportional Hazard Regression)라고 부르는 회귀분석 모형을 이용하고 있는 바, 유지보수 이력 데이터로부터 생존확률 초기값을 계산하고, 이 값은 다시 공정 변수들에 의해 보정이 됨



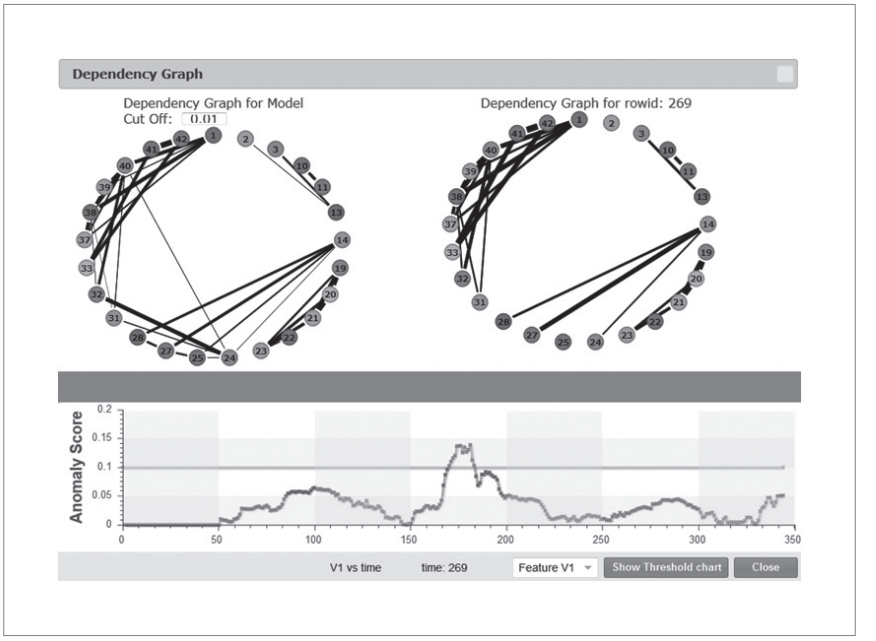
| Figure 10 고장위험분석의 출력물 예시 |

6-4. 이상거동 검출(Anomaly Detection)

이상거동 검출이란 정상적인 운영상태와 달라 보이는 예외적 거동을 찾아내는 작업임. 일반적으로 이상거동 검출에는 세가지 요소가 있음. 첫째는 정상상태의 모형임. 오늘날 분석시스템들은 실제 응용을 위한 정상상태의 모형으로 확률분포를 사용하고 있음. 둘째는 이상거동 점수의 정의로, 일반적으로 정상상태를 나타내는 확률분포의 함수로 표시됨. 셋째는 지금의 관찰결과가 이상거동인지 여부의 결정을 위한 이상거동 점수의 기준치임. 이상거동 검출에 기반해서 수행할 수 있는 몇 가지 파생 업무 중에서 변화감지는 IoT 센서에게 특히 현실적으로 중요성이 높음. 변화감지 과정에서, 일반적으로 슬라이딩 윈도 기술이 과거와 현재의 상태를 정의하는데 이용되고, 현재 상태와 과거의 차이를 표시하는 점수가 계산됨

적용사례 : 광산기계 이상거동 검출

★ 이는 지하 석탄광의 채굴에 사용되는 전단기(Shearer) 같은 고가의 광산기계에서 이상거동 상황을 검출한 사례임. 이러한 채굴 작업에서 광산기계에서 수집되는 센서 데이터는 매우 동적이어서 기존의 이상거동 검출 기법들은 적용에 한계가 있음. 추적이 어려운 특성 때문에 감시 솔루션들이 잘못된 경보를 남발하는 문제로 골머리를 앓고 있음. 이러한 문제의 해결을 위하여 이 사례에서는 기계학습 기법을 이용하여 Figure 11에 표시한 것 같이 이상거동 검출 임무를 변수간 의존성 구조의 변화를 정량화하는 것으로 정의하였음. 여기서 이상거동 점수의 최대값은 변수40과 관계가 없어진 변수24의 구조변화에 기인하는 것으로 나타남



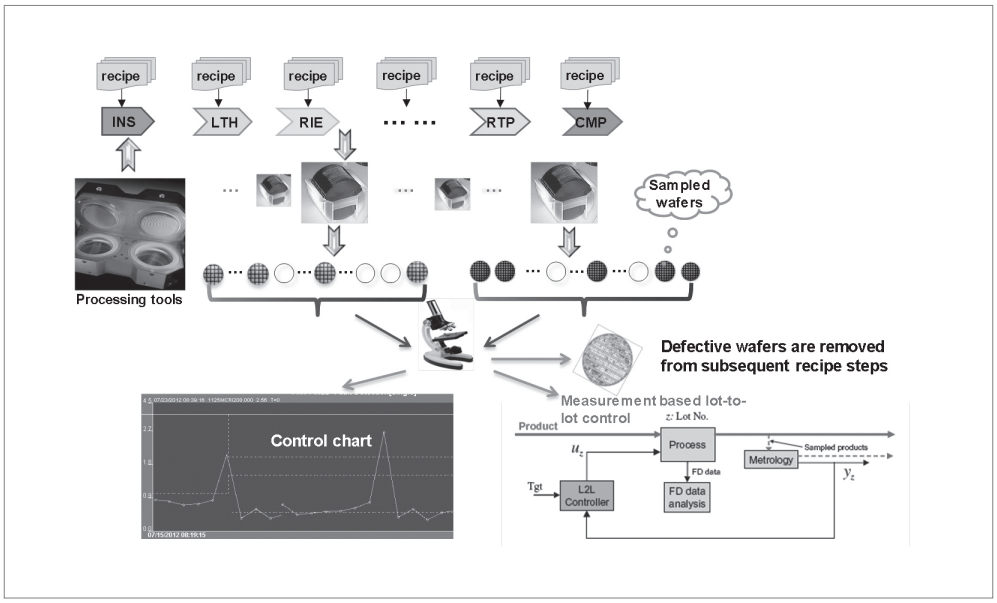
| Figure 11 고장위험도 분석시스템의 출력 예시 |

6-5. 예지적 통계적 품질/공정관리

예지적 통계적 품질/공정관리는 다양한 공정변수를 이용하여 공정 출력의 품질을 예측하는 모형을 개발하고, 향상된 피드백 제어를 위해 예측치를 활용하는 고급 회귀모형과 기계학습 접근방법의 하나임. 일반적인 기법으로는 릿지회귀(Ridge Regression), 라쏘회귀(Lasso Regression), 맞춤형 예측모형, EWMA, 인과관계모형, 신경회로망, 계층적 모형 같은 기법들이 있음

적용사례 : 반도체 제조의 가상 계측(Virtual Metrology)

★ 이 사례는 결과물의 실제 측정 데이터가 없는 상태에서 모형을 기반으로 공정 결과의 예측능력을 구축한 것임. 바탕이 된 모형은 실제 물리적 결과물의 이력과 공정 추적 데이터로부터 학습을 통해 제작된 것임. 시스템은 불량 제품의 조기 검출, 공정관리의 개선, 그리고 공정 감시 및 제어를 위한 물리적 측정의 필요성을 줄이는 데 도움이 되었음. 이 사례는 반도체 팹 공정에서 플라즈마 산화물 증착률의 예측에 사용되었음



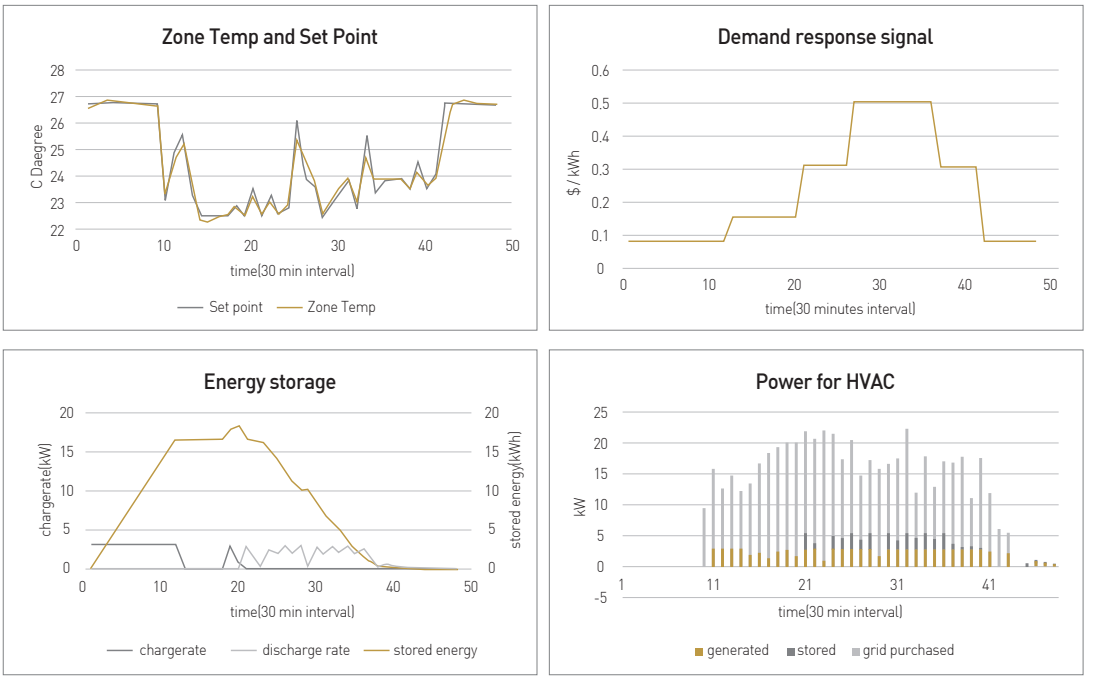
| Figure 12 예지적 통계적 품질/공정관리 분석시스템의 예시 |

6-6. 모형 예측제어(Model Predictive Control)

모형 예측제어(MPC: Model Predictive Control)는 석유-화학 플랜트나 발전시스템 같은 연속공정에 활용되는 최적 공정관리 기법임. MPC는 동적 공정모형, 특히 동적인 센싱 데이터를 통해 보정되는 물리학적 모형이나 데이터-구동형 모형을 이용하고 있음. 최적 제어문제는 통상 非선형계획(NLP) 모형으로 해결할 수 있음. MPC는 미래의 사건을 예측하고 이에 따른 최적제어 프로필(공정 최적제어를 위한 미래 일정기간 동안 제어변수 값들의 이력)을 계산함

적용사례 : HVAC 시스템의 최적제어

이는 수요에 따라 HVAC(Heating Ventilation And Condition: 냉난방 및 환기) 시스템을 최적으로 제어하는 MPC모형 적용사례임. 건물 구역의 열 거동(Thermal Behavior)을 NARNET(Non-Linear Autoregressive Neural Network: 非선형-자동회귀 신경회로망)으로 모형화 하고, 최적제어 문제는 MINLP(Mixed Integer Non-Linear Programming: 혼합정수 非선형계획법) 모형으로 정형화 했음. 이 최적제어 모형은 HVAC 장비와 현장의 에너지 저장 및 생성시스템의 물리적 한계 범위 안에서 건물 사용자들의 열적 쾌적성 수요를 충족하면서, 동적 수요반응과 현장의 에너지 저장 및 생산시스템 등을 고려하여 HVAC 시스템의 가동비와 온실가스(GHG: Green House Gas) 배출 규제에 따른 비용(탄소세 등)의 합을 최소화 하는 솔루션을 제공함



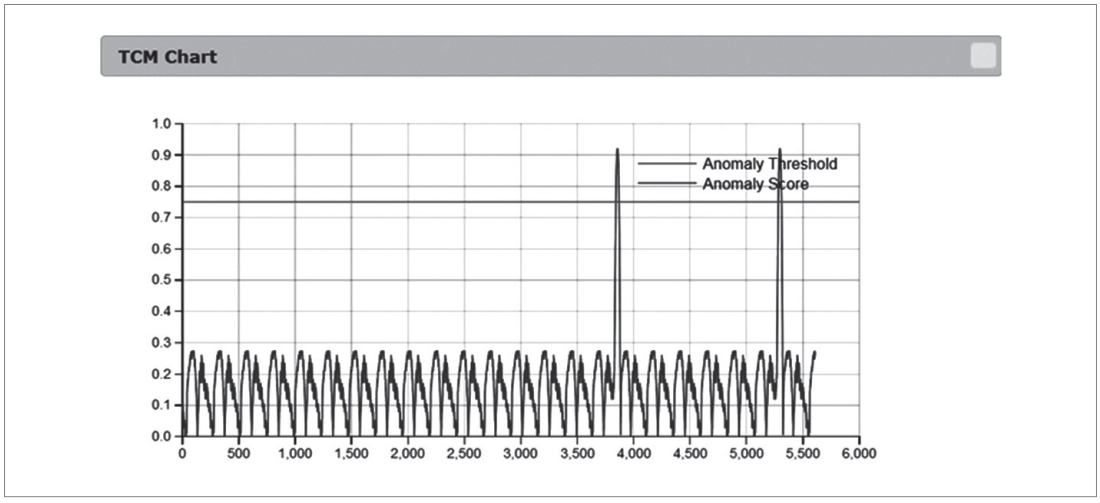
| Figure 13 모형 예측제어 분석시스템의 출력물 예시 |

6-7. 인과관계 모형

인과관계 모형은 현실세계 역학관계의 원인과 결과를 정량화한 모형임. 하나의 인과관계 모형은 다수의 변수 간에 존재하는 원인과 결과 및 기타 관계를 설명함. 인과관계 모형은 다중적 인과관계, 즉 어떤 결과이든 그 원인은 여러 가지가 있을 수 있다는 생각을 반영한 것임

적용사례 : IT 운영의 인과관계 분석

이는 컴퓨터 통신망과 시스템에서 발생하는 이상거동 상황을 감시 · 검출 · 진단하는 시스템 의 개발사례임. IT 데이터 모니터링 기술의 보급 확산과 분석대상 IT 운영 데이터가 홍수를 이루면서, 분석시스템에 기반한 운영에 따른 작업 효율 증대나 발생 문제의 이해 같은 혜택에 대한 기대치가 증가하고 있음. 분석시스템을 기반으로 관리하게 되면 특정 문제가 관련 서비스에 지장을 초래하기 전에 징후를 미리 감지해 낼 수 있음. 또한 기준에서 벗어나는 이상거동의 감지를 지원할 뿐 아니라, 근본원인 분석을 통해 능동적으로 고장을 예방할 수 있게 해 줄 것으로 기대됨



| Figure 14 SROM 분석시스템 플랫폼과 IoT/빅데이터 기반의 산업용 솔루션 |

7. 결론

★ 분석시스템은 산업의 새로운 천연자원인 데이터를 양질의 제품과 서비스로 변환하는 엔진임. 분석시스템이 없으면 우리가 수집하는 막대한 양의 데이터는 그저 데이터, 즉 잘 활용되지 못하는 천연자원일 뿐임. 데이터의 적절한 활용을 통하여 제조기업들은 제품 품질을 향상시키고, 원가를 절감하며, 공장 내 안전도를 향상시키고, 에너지 비용을 줄임으로써, 이익률을 향상시킬 수 있음. 이같이 분석시스템은 기업에 경쟁 우위를 제공할 수 있음

★ 또한 데이터도 스마트 건물, 스마트 도시, 스마트 지구의 구현을 촉진함으로써 삶의 질을 향상시키는 데 활용될 수 있음. 새로운 컴퓨팅 시대에 분석시스템은 선진적인 수학, 통계학, 데이터 마이닝, 기계학습, 인지적 컴퓨팅 및 최적화 등 다양한 알고리즘과 기법들이 복합된 집합체임. 따라서 산업계가 분석시스템을 직접 개발하여 활용하려면 많은 전문성과 노력 및 시간의 투자가 요구됨. 이에 기업들이 자원, 전문성, 시간의 큰 투자 없이 손쉽게 분석시스템을 사용할 수 있도록 지원하는 분석시스템과 이를 지원하는 서비스 라이브러리에 대한 수요가 있음. 여기에서 제시된 산업 범용 자원 및 운영관리 분석시스템 라이브러리는 그 엔진의 점화를 위한 것임

★ 새로운 컴퓨팅의 시대에 데이터는 가장 중요한 천연자원으로 자리매김 되었음. 천연자원은 정련하여 용도에 맞게 잘 가공해야 가치를 발휘할 수 있음. 데이터가 기대대로 새로운 부가가치를 창출하고 성장동력의 일익을 담당할 수 있으려면 데이터의 수집 · 정비 · 분석이 적절히 이루어져 최종적으로는 사용자의 최적 의사결정을 지원할 수 있어야 함. 이를 위해 우선 요인분석(Factor Analysis)을 통해 데이터의 사용 목적에 맞춰 필요한 데이터의 종류와 수량을 미리 이해하는 게 중요함. 이를 통해 필요한 데이터는 반드시 확보하되, 데이터 수집을 위한 정보화 인프라 구축 시

과잉투자를 방지할 수 있음. 이렇게 유용한 데이터의 확보기반 위에 영역별로 전문화된 분석시스템의 개발이 필수적임. 이러한 관점에서 여기에서는 제조산업의 자원 및 운영관리를 위한 분석시스템의 구조와 기능을 제시했음. 그러나 이 같은 접근방법은 제조산업뿐 아니라 마케팅 리서치, 에너지, 의료, 유통/물류, 국방, 레저, 문화, 교육, 연구개발 등 사회 모든 영역에 대해서도 유효한 것으로, 각기 특화된 전문적 분석시스템의 개발과 보급이 이어질 것으로 예상됨

(참고문헌)

1. IDC's sixth annual study of the digital universe, 2012. http://www.emc.com/leadership/digital-universe/2012iview/big-data-2020.htm
2. Big Data, Data savvy. Insight driven. http://www.ibm.com/big-data/us/en/
3. IBM Redbook, Performance and Capacity Implications for Big Data, http://www.redbooks.ibm.com/redpapers/pdfs/redp5070.pdf
4. IBM Redbook on The Interconnecting of Everything, http://www.redbooks.ibm.com/redpapers/pdfs/redp4975.pdf99
5. ITU Internet Reports 2005: The Internet of Things: Executive Summary: http://www.itu.int/osg/spu/publications/internetofthings/InternetofThings_summary.pdf
6. Big Data at the Speed of Business: What is big data: http://www.ibm.com/software/data/bigdata.