실무에 적용 가능한 Big Data 분석 개론

빅데이터 활용 사례



1. 빅데이터 활용 선진 사례 : 행복한 사회 구축

미국 국세청

탈세 방지 시스템 통한 국가 재정 강화

- 빅데이터 분석을 활용하여 탈세 및 사기 범죄 예방 시스템 구축
- 향후 발생할 수 있는 사기 범죄 및 탈세 관련 사건을 미연에 방지
- 세금 누락 및 불필요한 세금 환급 절감의 효과

일본

센서 데이터를 활용한 지능형 교통안내 시스템

- 실시간으로 GPS 데이터를 분석하여 최적의 교통 정보를 전달하는 서비스
- 출발지에서 목적지까지 최적 경로 안내
- 에너지 효율을 높이고 교통 체증 감소의 효과

1. 빅데이터 활용 선진 사례 : 행복한 사회 구축

이탈리아 밀라노

지능형 교통정보 시스템으로 길안내서비스 제공

- 밀라노 시내의 교통량과 속도를 조사하여 지능화된 교통정보 제공
 - 일정 기간 동안의 교통량, 속도, 기후 조건 등의 데이터를 종합 분석
 - 신속·정확하고 손쉬운 길안내서비스

뉴욕주 시라큐스시

데이터 분석을 기반으로 스마트 시티 추진

- 새로운 도시계획 수립 시 낙후 지역 활성화를 위해 데이터 분석 추진
 - 주민들에게 필요한 사항을 적극 고려함으로써 과학적인 정책 추진을 위한 기반 마련

1. 빅데이터 활용 선진 사례 : 행복한 사회 구축

덴마크 베스타드 윈드 시스템

풍력 에너지 관리로 에너지 생산 효과 극대화

- 베스타드 윈드 시스템: 덴마크 풍력 발전회사
 - 데이터 활용으로 풍력 발전에 필요한 터빈 관리 및 배치를 효율적으로 추진

구글

실시간 자동 번역시스템을 통한 의사소통 불편 해소

- 통계학적으로 가장 정확한 번역 표현 검색
 - 기존 방법과 달리 전문가가 번역한 문서를 데이터베이스화해서 비슷한 문장과 어구를 대응시키는 통계적 기법 활용

1. 빅데이터 활용 선진 사례 : 행복한 사회 구축

월마트

데이터 분석을 통한 투자수익 증대

- 각 지점의 모바일과 소셜 쇼핑의 특징을 이용한 월마트랩 운영
 - 거래 데이터를 이용한 재고 예측 조사 시스템 마련
 - 고객 선호도 및 수요 데이터 분석을 통해 더 나은 비즈니스 환경 구축

자라

점포별·상품별 등 실시간 데이터 분석을 통한 판매량 증대

- 실시간으로 수집한 매장 판매 데이터를 통해 최대 매출이 가능한 재고 최적 분배 시스템 활용
 - 불필요한 재고 관리 감소
 - 실시간으로 전달된 고객 니즈가 반영된 제품 생산 및 관리 가능

1. 빅데이터 활용 선진 사례 : 행복한 사회 구축

마이크론 테크놀로지

제품생산시간 분석을 통한 비용절감

- 결과에 영향을 미치는 모든 요소를 통합 분석
 - 작업시간 절감 및 업무 효율 증대
 - 제품생산과 관계된 의사결정에 필요한 과학적 근거 마련

코카콜라

SNS 데이터 활용을 통한 제품 판매 의사결정 반영

- 코카콜라에 비우호적인 정보가 증가하는 국가나 지역을 대상으로 홍보를 강화하는 등 실시간 대응 가능
- 이용자들의 호불호가 빠르게 반영되는 SNS 데이터를 서비스 개선에 실시간으로 적용하여 고객 맞춤형 서비스 제공 가능

1. 빅데이터 활용 선진 사례 : 행복한 사회 구축

리츠칼튼 호텔

데이터 관리를 통한 고객맞춤형 서비스 제공

- 고객 데이터의 수집과 과학적인 분석을 통한 고객니즈 파악
 - 전 세계 고객 100만 명의 데이터베이스 공유
 - 고객의 요구사항을 알아서 파악하고 그에 맞는 서비스를 제공하는 지능형 친절 시스템

할리우드 영화시장

SNS 분석을 통해 흥행 수익 예측

- 실시간으로 사용자의 의견 분석 가능
- 다양한 사람의 의견을 수렴하는 용도로 발전 가능
- 각 분야에 대한 의견 분석 도구로 발전 예상

1. 빅데이터 활용 선진 사례: 행복한 사회 구축

넷플릭스

데이터 분석으로 온라인 DVD 판매제고 및 고객 서비스 향상

- 데이터 분석을 통해 DVD업계 최고 기업으로 부상
 - 넷플릭스 : 소규모 비디오 렌탈 업체에서 콘텐츠 미디어 유통기업으로 급속 성장
 - 고객들의 영화 취향 데이터를 분석하고 고객이 가장 선호할 영화를 추천하는 시스템 개발

2. 미국 국세청의 탈세 방지 시스템 개발을 통한 국가 재정 강화 사례

추진목적 및 배경

- 💶 탈세 및 사기로 인한 국가 재정 위기의 가능성이 증가하고 있음
- 글로벌 금융위기로 인한 재정 위기의 확산은 개인과 기업의 탈세에 따른 낭비성 재정 지출 문제를 발생하게 함
 - ▶ 세계은행은 탈세 및 세금 사기와 관련된 지구촌 지하경제의 규모가 전체 GDP의 18%에 이를 것으로 전망함
 - ▶ OECD는 지하경제에 흘러 다니는 탈세를 조세망으로 끌어들여 지속 가능한 경제 발전을 이루어야 한다고 경고함
- 3 전세계적으로 탈세 금액이 증가하고 있음
 - ▶ 미국의 탈세 금액은 2010년을 기준으로 저속득층 의료보장 총액을 초과하여 GDP대비 8.6% 정도 수준으로 추정하고 있음

추진내용

▶ 탈세 및 사기 방지 시스템

방대한 자료로부터 사기관련 이상 징후를 찾아냄

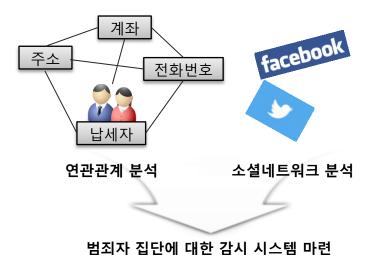
사기 예측 모델링을 통해 과거의 행동 정보 분석

사기 패턴과 유사한 행동을 검출하는 시스템을 개발

2. 미국 국세청의 탈세 방지 시스템 개발을 통한 국가 재정 강화 사례

추진내용

▶ 범죄 네트워크 발굴



▶ 지능형 감시 시스템 구축

데이터베이스를 통합해 대용량 데이터를 효율적으로 활용하고 프로세스를 통해 지능형 데이터 분석 능력을 지원할 수 있는 시스템



2. 미국 국세청의 탈세 방지 시스템 개발을 통한 국가 재정 강화 사례

효과

- 통합형 탈세 및 정부사기 방지 시스템을 통해 연간 3,450억 달러에 달하는 세금 누락 및 불필요한 세금 환급 절감
- 과학적 데이터를 근거로 탈세 조사를 수행함으로써 탈세자 수의 감축 등의 성과 발생
- 과거 데이터 분석을 통해 향후 발생할 수 있는 사기 범죄 및 탈세 관련 사건을 미연에 방지

정책적 시사점

- 사기, 재정낭비, 부당 지출과 관련된 여러 첨단 기술을 접목시켜 다양한 정부 사업에 응용 가능
- ② 의료보험 및 복지프로그램 관련 사업에도 동일한 기술의 응용 가능

실무에 적용 가능한 Big Data 분석 개론

빅데이터 활용 사례



1. 빅데이터 활용 선진 사례 : 건강한 사회 구축

미국 국립보건원(NIH)

유전자 데이터 공유를 통한 질병치료체계 마련

- 다양한 질병을 연구하기 위해 유전자 데이터를 공유 및 분석할 수 있는 시스템 마련
 - 주요 관리 대상에 해당하는 질병에 대한 관리 및 예측

Pillbox 프로젝트를 통한 의료개혁

- 미국 국립보건원 산하 국립의학도서관(National Library of Medicine)이 운영하는 약 검색 사이트
 - 약 검색 서비스를 통해 얻어진 다양한 사용자의 질병에 대한 통계 데이터 활용
 - 주요 관리 대상에 해당하는 질병에 대해 국가차원의 조기 대응이 가능

미국

퇴역군인의 <mark>전자의료기록 분석을 통한 맞춤형 의료 서</mark>비스 지원

- 퇴역군인의 의료 기록을 분석하여 고품질 의료 서비스 지원
 - 퇴역군인의 의료 기록을 통해 수집된 데이터로 효과적인 의료 지원 가능
 - 의사와 환자간의 상호작용이 매우 원활할 것을 전망

1. 빅데이터 활용 선진 사례: 건강한 사회 구축

싱가포르

주민위원회 센터 네트워크를 기반으로 맞춤형 복지사회 구현

- 데이터 분석을 통한 지역사회의 화합 및 맞춤형 복지사회 구현
 - 다양한 인종, 나이, 문화, 소득, 연령에 따른 주민의 데이터를 수집· 분석하여 맞춤형 서비스 제공 가능한 플랫폼 구현

캐나다 온타리오 공과대병원

미숙아 모니터링을 통한 감염 예방 및 예측

- 인큐베이터 내 미숙아의 다양한 데이터를 분석하여 병원균 감염을 예측할 수 있는 시스템
 - 의사전달이 힘든 환자를 위한 진단 및 치료 시스템으로 발전 가능

1. 빅데이터 활용 선진 사례 : 건강한 사회 구축

건강보험회사 웰포인트

슈퍼컴퓨터를 활용한 효율적인 환자치료

- 의료진의 진단과 환자 치료에 슈퍼컴퓨터를 사용하는 시스템
 - 과거 데이터를 분석하여 진단율 및 치료율 향상
 - 적절한 치료법 제공 및 최신 정보를 과학적인 방법으로 제시 가능

구글

검색어 분석을 통한 독감예보 서비스 제공

- '감기'와 관련된 검색어 분석을 통한 독감 예보 시스템 제공
 - 구글 검색 사이트에 사용자가 남긴 검색어의 빈도를 조사, 독감 환자의 분포 및 확산 정보 제공
 - 실시간으로 사용자 의도를 분석할 수 있는 기술 개발로 더 빠르고 정확한 통계의 확보 가능

1. 빅데이터 활용 선진 사례: 건강한 사회 구축

네덜란드 스파크드

빅데이터를 활용하여 건강한 소 사육 환경 구축

- 소 사육에 필요한 소의 움직임, 건강, 기후 등의 데이터를 종합적으로 분석
 - 효율적인 소 사육을 통해 궁극적으로 축산 산업 발전에 기여
 - 객관적 데이터를 기반으로 농장 위생환경 개선 및 축산 산업 경쟁력 제고

2. 미국 국립보건원의 유전자 데이터 공유를 통한 질병치료체계 마련 사례

추진목적 및 배경

- 1
- 인종별, 국가별 특성에 따른 유전자 특성 정보를 분석하여 질병 치료에 활용하고자 함
- ▶ 인간의 35,000개 유전자에 따라 같은 질환이라도 증상이 다르게 나타나며, 약물반응도 다르게 발생
- ▶ 미국에서는 한 해 220만 명 이상이 약물 부작용에 의해 입원하고 있고, 약물 부작용은 미국인 사망원인 4위일 정도로 심각함
- ② 인간 유전체의 다양성을 알아내고자 함
 - ▶ 개인의 유전체는 유전자의 기능이 무력화된 250개에서 300개의 기능을 잃은 돌연변이에 의해 발생함
 - ▶ 발견된 돌연변이들 중 50%는 향후 개인별 유전체를 연구할 때 큰 도움이 될 것으로 전망됨

추진내용

200TB의 유전자 데이터 확보

아마존 웹 서비스에 저장

아마존 클라우드에 저장

1000 유전체 프로젝트(1000, Genomes Project)



200TB의 유전자 데이터 확보

전세계에서 2,662명의 유전자 정보 저장

1% 이상의 빈도를 나타내는 유전적 다양성 분석

2. 미국 국립보건원의 유전자 데이터 공유를 통한 질병치료체계 마련 사례

추진내용



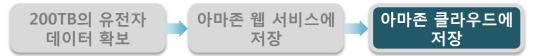
아마존 웹 서비스(AWS, Amazon Web Service)



<출처: The 1000 Genomes Tutorial The Website and Browser, 2012.2.17>

2. 미국 국립보건원의 유전자 데이터 공유를 통한 질병치료체계 마련 사례

추진내용



1,700명의 유전자 정보를 아마존 클라우드에 저장



아마존 웹 서비스에 유전자 정보를 넣으면 아마존 웹 서비스 서버에 저장됨

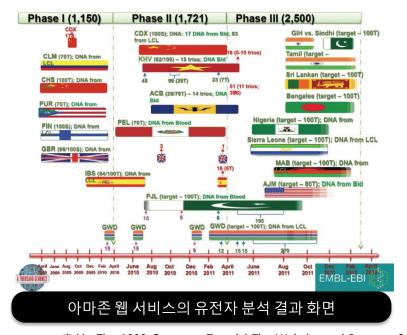


1,700개 이상의 유전자 정보는 익명의 개인들로부터 취합



전세계 인구 표본에서 데이터를 취합함

- ▶ 북동유럽에서 조상들이 미국으로 이주하고 현재 유타주에 거주하는 사람
- ▶ 덴버에 사는 중국계
- ▶ 로스앤젤레스에 사는 멕시코계
- ▶ 미국 남서부에 거주하는 아프리카계 등



<출처 : The 1000 Genomes Tutorial The Website and Browser, 2012.2.17>

2. 미국 국립보건원의 유전자 데이터 공유를 통한 질병치료체계 마련 사례

효과

- 유전자 정보를 공유함으로써 새로운 질병에 대한 **빠른 진단 서비스 제공**이 가능함
- 안치병 및 불치병과 관련된 유전자 정보를 공유하고 분석함으로써 새로운 치료제 개발 가능성을 높임
- 최신 IT 기술을 결합함으로써 기존에 치료하기 힘들거나 오래 걸리는 질병에 대한 <mark>치료 확률을 높일 수 있는 기회</mark>를 마련함

실무에 적용 가능한 Big Data 분석 개론

빅데이터 활용 사례



1. 빅데이터 활용 선진 사례 : 안전한 사회 구축

싱가포르

국가 위험관리 시스템을 통한 국가안전관리

- 국가적 차원의 위험요인과 기회요인을 선제적으로 파악, 대응 방안을 수립하기 위한 시스템
 - 국가와 국민의 위험 요소를 사전에 예측하고 대응 방안을 모색하기 위한 과학적 시스템 마련

FBI

유전자 색인 시스템 활용한 단시간 범인 검거 체계 마련

- 유전자 분석표를 대조함으로써 사건을 해결
 - 과거 범죄자들의 유전자 정보를 데이터베이스화하여 빠르고 과학적인 수사가 가능
 - 민감하고 소중한 개인정보의 오남용 우려 존재

1. 빅데이터 활용 선진 사례 : 안전한 사회 구축

샌프란시스코

범죄 예방 시스템으로 안전 지역사회 구축

- 범죄 발생 지역 및 시각을 예측하여 범죄를 미연에 방지하기 위한 시스템
 - 과거 발생한 범죄 패턴을 분석하여 후속 범죄 가능성 예측
 - 과거 데이터에서 범죄자 행동을 분석하여 사건 예방을 위한 해법 제시

싱가포르 출입국관리소

통합적 정보분석으로 출입국 보안 및 국경 통제 강화

- 출입국자, 수화물, 운송수단에 대한 통합적 정보 분석으로 위험 요소에 대한 효과적 관리
 - 사용자가 편리한 출입국 서비스 제공
 - 위험 입국자를 보다 신속, 정확하게 파악

1. 빅데이터 활용 선진 사례 : 안전한 사회 구축

일본

다양한 센서 데이터를 활용한 재난대응 능력 강화

- 4세대 이동 통신 기술, 인공위성망 등을 통하여 현장 대응 인프라를 강화
- 화재, 방사능 오염 등 재난 현장의 데이터 분석
- 사람이 접근하기 힘든 재난현장에 대한 상황 파악 및 현장 대응 가능

- 빅데이터 활용을 통한 안전한 사회 구축 사례
 - 2. 싱가포르의 국가 위험관리 시스템을 통한 국가안전관리 사례

추진목적 및 배경

- 1
- 9.11 이후 국제 사회는 테러와의 전쟁을 전개해 오고 있으나 테러의 공포로부터 벗어나지 못하고 있음
- ▶ 국내 여러 계층간, 국가간 또는 종교적인 갈등에 의한 테러 전개
- ▶ 민간인을 대상으로 한 무차별적 테러로 인해 국민들의 불안감 증대
- 2
- 재난, 전염병 등 국민의 안전과 건강을 위협하는 요소에 대한 사전 감지 및 예방이 요구됨
- ▶ 환경안보, 해상안보, 사이버안보, 식량안보 등 다양한 재난 상황 시뮬레이션
- ▶ 상황별 대응조치 마련 및 위기 시 <mark>즉각적으로 대응하는</mark> 체계적 지원체계의 수립
- 3
- 유로존 국가의 위기가 전 세계 금융시장으로 확대되는 동시에 실물 경제로의 전이가 본격화됨
- ▶ 유로존 국가 및 금융기관 간의 상호 연계성이 깊기 때문임

2. 싱가포르의 국가 위험관리 시스템을 통한 국가안전관리 사례

추진내용

성가포르의 국가 위험관리 시스템 (RAHS, Risk Assessment and Horizon Scanning)

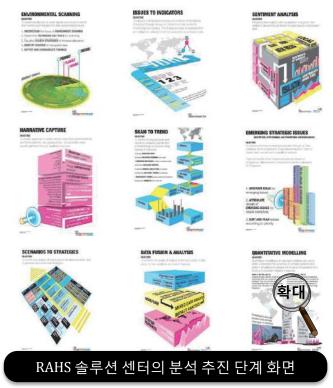


질병, 금융 위기 등 모든 국가적 위험을 수집 및 분석



불확실한 미래를 대비하기 위하여, 2004년부터 빅데이터를 기반으로 한 위험관리 계획을 추진함

수집된 정보는 시뮬레이션, 시나리오 기법을 통해 분석하여 사전 위험 예측 및 대응방안 모색에 활용함



<출처 : www.hsc.gov.sg>

2. 싱가포르의 국가 위험관리 시스템을 통한 국가안전관리 사례

효과

- 국가 및 국민의 위험 요소를 파악하고 대비함으로써 국민의 생명과 재산을 보호할 수 있음
- 수많은 데이터와 변수를 동시에 고려하는 전천후 국가 위험 관리 체계로 발전할 수 있음
- 정부가 보유한 다양한 정보를 체계적으로 관리 및 분석함으로써 효율적인 위험 관리 시스템으로 발전할 수 있음
 - ▶ 국가의 위험 요소는 국내적 요소 외에 국가간 발생 요소들이 더 중요하므로 다양한 변수를 고려하는 것이 필수적임

실무에 적용 가능한 Big Data 분석 개론

빅데이터 활용 사례



1. 빅데이터 활용 선진 사례 : 창의적 사회 구축

미국 미시간 주

데이터웨어하우스 구축으로 공공서비스 질적 향상

- 미시간 주의 21개 정부기관 데이터 통합 운영
- 정보의 효율적 이용 가능
- 시민에게 보다 나은 정보 제공 가능

영국 패치베이

국민참여형 안전관리 플랫폼 구현

- 공공 · 민간 · 개인 등이 보유한 정보를 개방하여 상호 공유하게 하는 서비스
 - 공유된 데이터를 웹 프로그램, 스마트폰 앱 개발 등에 활용 가능

1. 빅데이터 활용 선진 사례 : 창의적 사회 구축

케냐 우샤히디

집단지성으로 이루어진 재난관리 오픈소스 플랫폼

- 우샤히디(Ushahidi)
 - 세계 각국에서 발생하는 폭력, 테러, 자연재해 등을 실시간으로 정보 공유하는 시스템
 - 위치기반을 이용해서 자연재해, 선거부정, 범죄 등에 대응할 수 있는 IT기반 재난안전 플랫폼

IBM 왓슨

인공지능 슈퍼컴퓨터로 인류의 창조성과 혁신 촉진

- 인간의 언어를 이해할 수 있는 첫 번째 컴퓨터로 과학과 기술의 발전을 촉진
 - 다양한 분야에 활용가능함

1. 빅데이터 활용 선진 사례 : 창의적 사회 구축

애플 시리

지능형 음성인식을 통해 더 똑똑해지는 창의적 사고 가능

- 사용자가 원하는 검색 결과나 기능을 실행해주는 서비스
 - 향후 시리와 상거래를 결합하면 엄청난 IT산업으로 성장할 가능성 존재

대형강입자충돌기를 이용한 힉스입자 검출

• 양성자 충돌을 통해 얻은 수많은 데이터를 분석하여 힉스입자에 대한 비밀을 밝히는데 기여

2. 영국 패치배이의 국민참여형 안전관리 창의 플랫폼 구현 사례

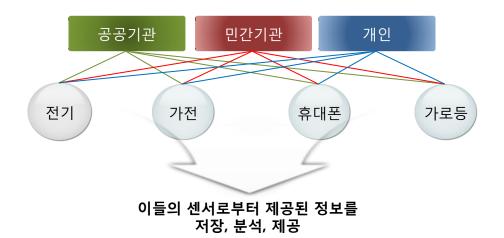
추진목적 및 배경

- 1 센서로부터 입력된 수많은 정보를 분석하여 데이터화할 필요가 있음
 - ▶ 유비쿼터스 환경 구현을 위한 기반 구조로써 WSN(Wireless Sensor Network) 기술 개발에 대한 관심 고조
 - ▶ 다양한 센서 기술과 네트워크 환경 발전에 따라 실시간으로 데이터 수집이 가능
- 공공부분, 산업, 개인의 데이터가 축적되면서 천문학적 수치의 데이터가 생성되고 있음
 - ▶ 2010년 100억 개의 데이터가 다양한 센서 및 디지털 기기로 수집되었다면, 2020년에는 7조 개로 증가할 것으로 예상
 - ▶ 2017년 세계 70억 인구가 7조 개의 무선 단말기로 연결될 것으로 예상

추진내용

수<mark>많은 센서로부터 입력된 전력, 환경 등의</mark> 정보를 개방, 공유하는 플랫폼 개발

2010년 영국에서 시작됨



2. 영국 패치배이의 국민참여형 안전관리 창의 플랫폼 구현 사례

추진내용



개방된 소스로서 재난 안전관리 시스템의 상호 연계 지원



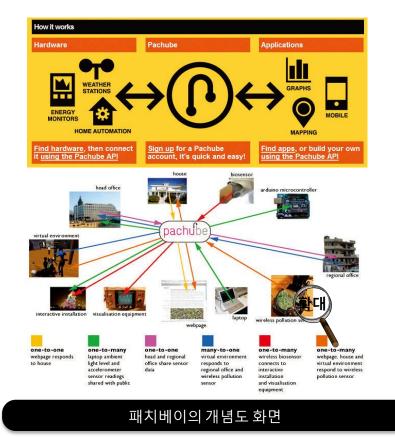
공유 데이터를 기반으로 웹 프로그램, 스마트폰 앱 개발 등에 응용 및 활용



일반적인 수준의 데이터 이용은 대중에게 무료로 공개함



앱 개발 또는 분석된 데이터의 독점적 이용을 원하는 고객들에게는 유료로 제공



<출처 : www.pachube.com

2. 영국 패치배이의 국민참여형 안전관리 창의 플랫폼 구현 사례

효과

- 다양한 센서로부터 얻어진 정보를 실시간으로 분석함으로써 원하는 정보를 빠르게 확보할 수 있음
- ② 공공기간, 개인, 민간이 가지고 있는 센서 데이터를 통합적으로 분석하여 각종 데이터를 효율적으로 이용할 수 있음
- 스마트폰에서 앱 등의 다양한 플랫폼을 제공하여 사용자가 원하는 정보에 손쉽게 접근하는 것이 가능함
- 공공, 민간기업, 개인의 센서 정보를 망라하기 때문에 정확한 정보의 수집이 가능함
- 개방형 플랫폼으로써 누구든지 정보를 가공하여 원하는 애플리케이션 형태로 정보를 제공하는 것이 가능함