

방대한 데이터에서 의미를 건져낸다는 것

박형주 국가수리과학연구소장

수와 예술의 탄생



수학

우아함과 대칭성의 추구?
or
인류에게 닥친 문제 해결?

어원

수학: 그리스어 mathema, ‘배우는 것 (to learn)’

과학: 라틴어 Regina Scientiarum, 지식 분야 (field of knowledge)

선사시대

BC 10만년

-- 호모 사피엔스의 출현

-- 언어의 발명

BC 5만년

←--- 수의 발명

BC 5천년

-- 문자의 발명

역사시대

예술의 탄생



빌렌도르프의
비너스

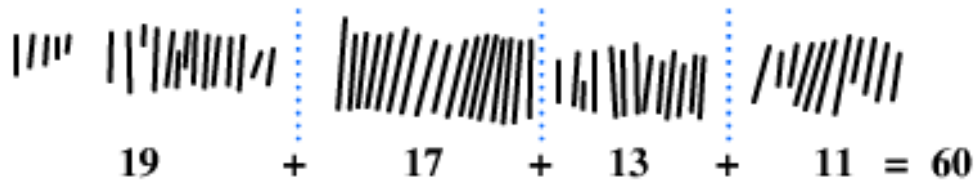
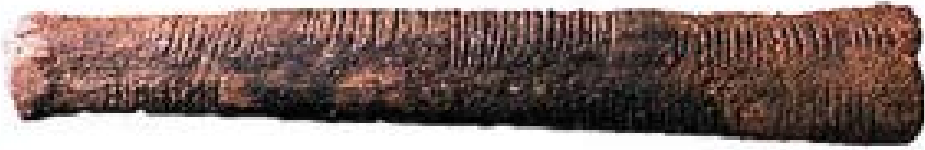
BC 23,000년



동물뼈 플루트

BC 35,000년

수학의 탄생



Ishango 뼈
BC 25,000년

- **문자의 발명 이전에** 생존의 필요로 셈(counting)의 개념이 출현했음을 입증
- 사냥감의 개수를 기록한 것이라는 견해와 수치 연산의 기록이라는 견해
- 달의 모양 변화를 기록한 것?
→ 기록자가 여성?

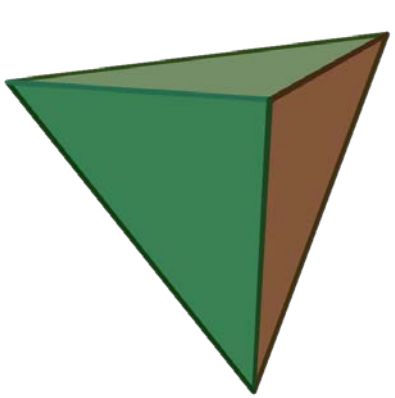


살바도르 달리 최후의 만찬 (1955)

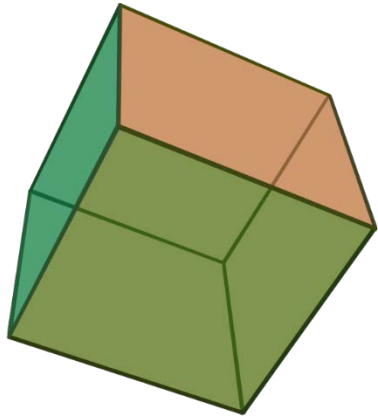
달리는 왜
정십이면체를 배경으로
사용했을까?

플라톤의 입체

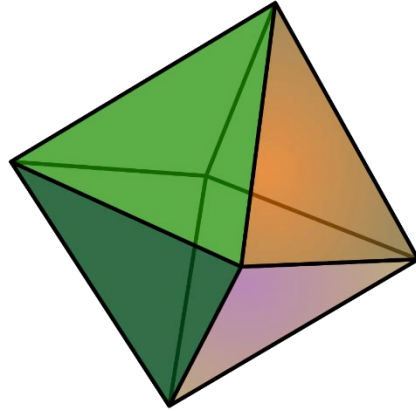
모든 면이 똑같은 정다면체는
우주에 다섯 개 밖에 없다



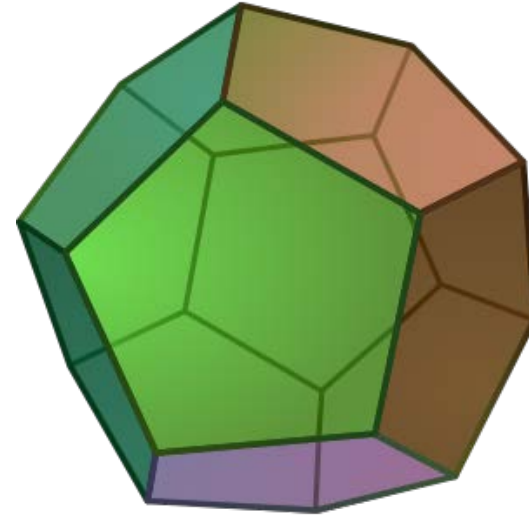
정사면체
→ 불



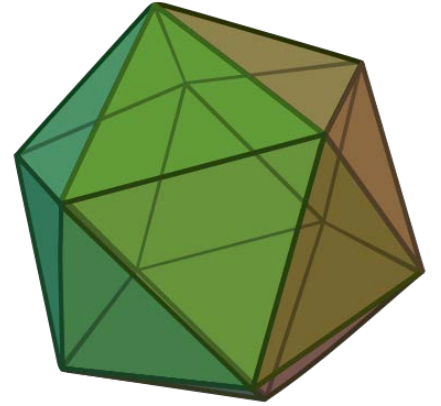
정육면체
→ 흙



정팔면체
→ 공기



정십이면체
→ 제5원소??



정이십면체
→ 물

플라톤은 제5원소를 신이 우주의 별자리를 배치하는 데 사용했을거라고 생각
아리스토텔레스는 에테르가 제 5원소라고 여겼다



이질적인 것들의 만남과 연결

**Uber와 Air B&B가
파격적인 기술력이나 특허를 가졌나?**

Zara의 옷이 정말 특별한가?

**Amazon의 물류와 배송을 혁신하고 9천억원에
매각된 Kiva 시스템즈는 넘사벽의 하드웨어
기술력을 가졌나?**

4차산업혁명의 핵심은

(새로운 기술의 개발보다는)

이미 존재하는 기술을 연결하는 것

기술이 최고라고?

기술적 우월성 → 경쟁자의 추격 → 오래 버티기 힘들다

높은 수준의 복합적인 문제 해결력이 주효한 분야는 지속적인 경쟁우위 유지 가능

기술보다는 문제해결능력이 미래 경쟁력

인공지능과 빅데이터 시대의 도래

빅데이터의 부상

- 인간 삶의 전 영역에서 방대한 데이터가 창출됨
 - ✓ “지난 2년 간 세계 휴대폰을 통해 창출된 데이터는 인류의 2천년 역사에서 창출된 데이터 총량과 비슷” - 2011년 메킨지 보고
- 방대한 데이터는 숫자의 나열에 불과. 그 속에 숨겨진 뜻을 이해하고 소통할 수 있는 능력이 일자리가 요구하는 주요 소양이 될 것
 - ✓ “수학적 이해도를 가진 데이터 분석가와 방대한 데이터에 의해 만들어지는 복잡한 상품을 이해하고 일반인의 눈높이에서 설명할 수 있는 전문 세일즈 요원 필요” - 2016년 다보스포럼 미래일자리보고서

패션산업의 예



ZARA

- 세계 최대의 의류 회사
- 인스턴트 패션과 소량다종 생산의 효율성으로 성장
- 창업자 아만시오 오르테가는 Forbes 세계 부자 순위에서 2016년 10월 기준 1위였고, 2017년 3월 기준으로 4위

지식의 시대 저물다



21세기가 지식의 시대가 아니라는 역설

새로운 지식이 너무 빨리 출현하니,

→ 오히려 얼마나 아는냐는 덜 중요

어차피 수년 내에 낡은 지식이 될 것이고

→ 새로운 걸 계속 배워나가야

대학에서 전공한 지식도

→ 직장에서 유용한 기간이 짧고, 곧 ‘옛’ 지식

21세기는
‘지식 과잉’ 과 ‘무한 정보’ 로
요약된다

방대한 지식과 데이터에 묻혀 길을 잃는 사람들



데이터의 함의를 읽어내며 닥친 문제의 본질을
보고 해결하는 능력이 시대를 이끈다

**바야흐로 지식의 시대는 저물고
통찰의 시대가 온 것이다**

과학 발전과 일자리 문제

과학기술의 발전과 산업혁명

18세기: 증기기관 $\Rightarrow$ 기계 생산 $\Rightarrow$ 1차 산업혁명

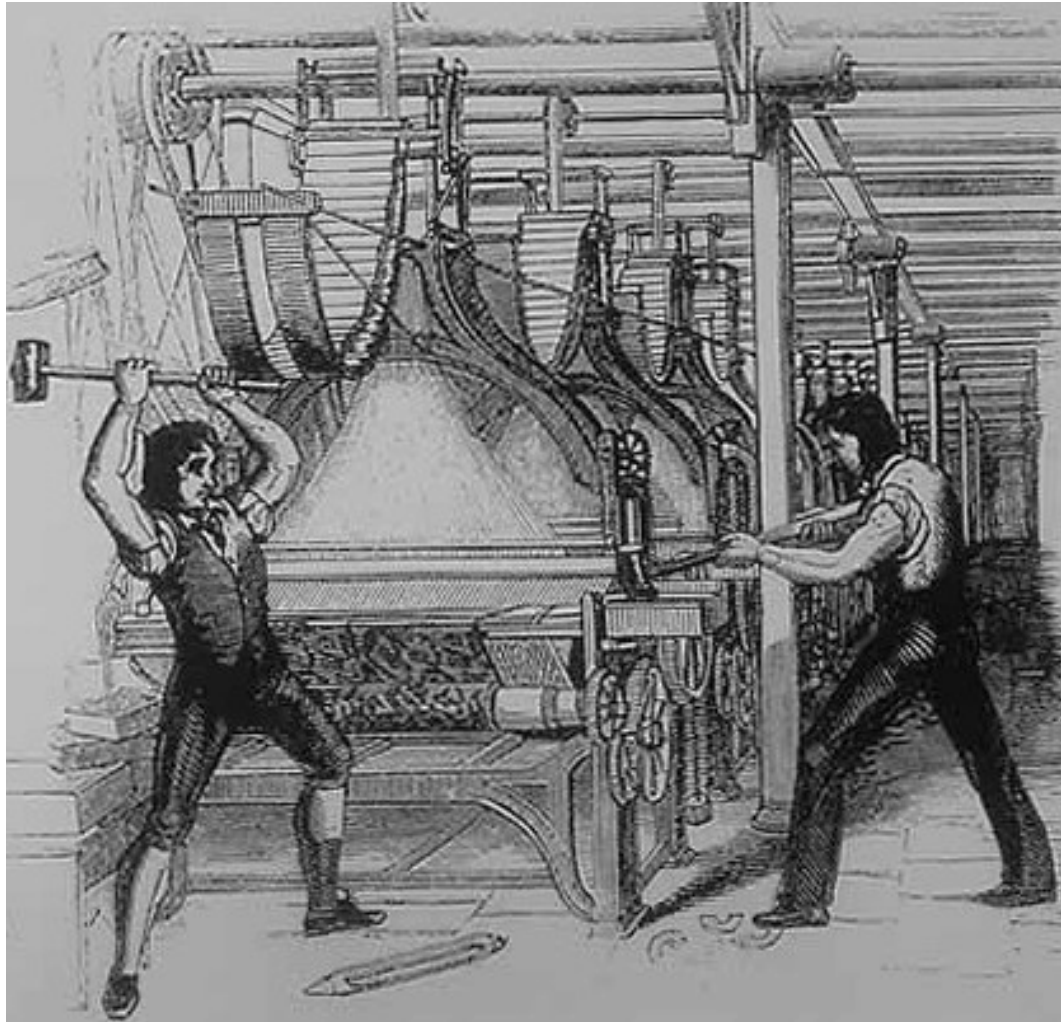
19세기~20세기 초: 전기에너지 $\Rightarrow$ 대량생산 $\Rightarrow$ 2차 산업혁명

20세기 후반: 컴퓨터 + 인터넷 $\Rightarrow$ 생산자동화 $\Rightarrow$ 3차 산업혁명

현재: 로봇 + 인공지능 $\Rightarrow$ 실세계 + 가상세계 $\Rightarrow$ 4차 산업혁명

산업혁명으로 일자리가 사라지자...

Luddite 운동



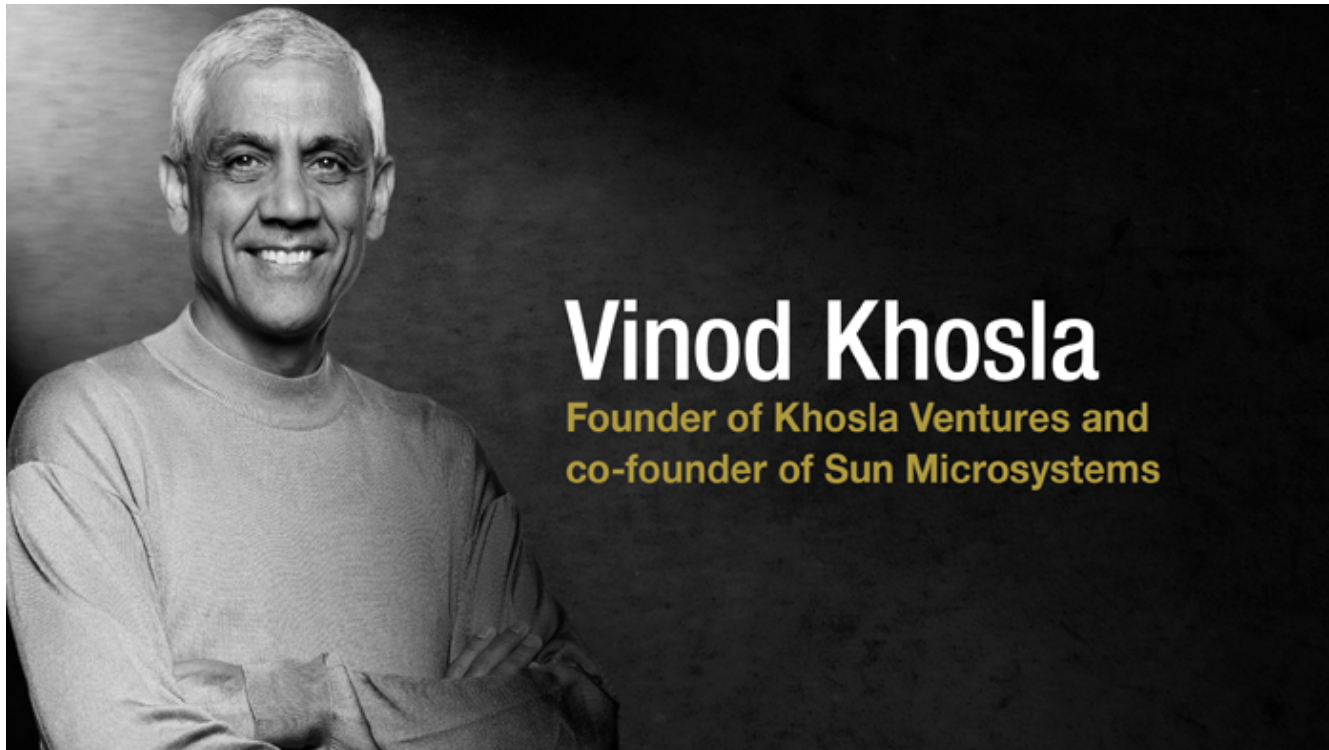
일자리 감소

- **로봇의 활용도 증대**
 - 전체 일자리의 절반 정도가 로봇으로 대체 가능
- **공유경제의 개념 등장**
 - 차량과 숙박업소 감소 및 종사자 일자리 사라짐
- **자율주행자동차**
 - 택시운전사와 버스운전사 일자리 사라질 가능성
- **핀테크**
 - 은행의 대면업무 감소
 - 2015년 미국과 유럽 은행원의 10%인 10만 명이 일자리 잃음.
10년 뒤 은행 관련 일자리의 절반 사라질 것으로 예상
- **빅데이터 무인진단법**
 - Vinod Khosla는 의사 수가 80% 격감할 것으로 추산



Vinod Khosla

“Machines will replace 80 percent of doctors”



- Sun Microsystems 공동창업자
- 실리콘 밸리의 주요 벤처 투자자

지능정보기술 & 수학

**Mobile First
To
AI first**

**Sundar Pichai
Google CEO**

**BSc in Artificial
Intelligence and
Mathematics**

**University of
Edinburgh**

인공지능 기술의 핵심

**‘방대한(너무 큰)’ 데이터와의 비교를 통해
최적의 solution을 찾는 것**

어떻게?

계산 자원

(딥블루의 체스 정복: 2억번의 수읽기)

또는

수학적 알고리즘

(알파고의 바둑 정복: 30초 동안 10만번의 수읽기)

(이창호 9단: 약 100수의 수읽기)

**알파고를 필두로 한 새로운 인공지능
기술의 핵심은
계산 자원의 문제가 아니라**

**혁신적인 수학 알고리즘의 도입을 통해
계산 자원의 필요를 획기적으로 줄이는
데 성공했다는 것**

수학 알고리즘이 방대한 빅데이터 문제의 해결책인 이유

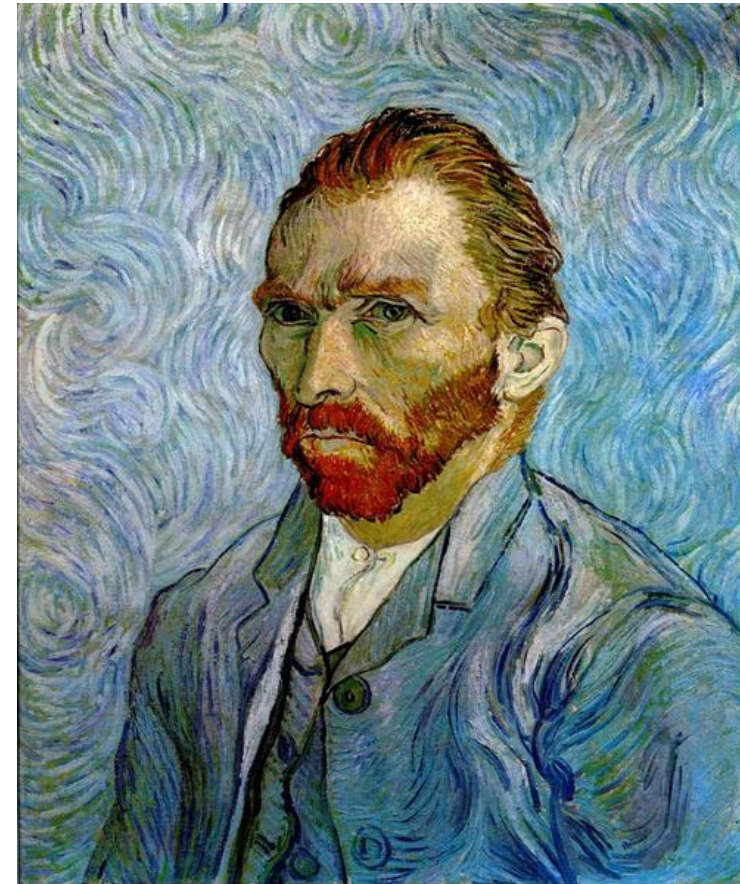
from **BIG** Data to **SMALL** Data

패턴 인식을 통해 불필요한 경우를 배제하여
소수의 데이터에 집중

거대 문제를 작은 규모 문제로 바꾸면서도 정보 손실을 최소화 하는 수학적
알고리즘이 인공지능의 핵심 기술로 부상

사례: 미술 위작품 감정

- 2008년. 미국의 방송제작업체 NOVA는 반 고흐의 작품 6개를 제시하고 이 중에 숨어있는 위작품 하나를 찾아내는 챌린지 진행. 참가팀들이 이에 도전하는 과정은 다큐로 제작되어서 PBS에서 방송
- 프린스턴 대학 수학자 잉그리드 도브시 교수 팀은 웨이블릿(wavelet)이라는 수학기론을 무기로 챌린지에 참가 성공적으로 위작을 가려냄
- **Main Idea:** 원작자는 자기 생각의 표현에 집중하지만, 위작자는 원작과 동일해야 한다는 강박관념에서 자유롭지 못함. 따라서 모사과정에서 구체적인 선과 곡선을 그려낼 때 눈에 보이지 않는 ‘주저함’이 숨어있을 거라고 추정. 이 주저함을 수학적으로 정량화해서 찾아냄
- ‘모방작에 숨어있는 주저함의 정도’를 추적 → 놀라운 관점의 전환



사례: FBI가 지문으로 범인 찾는 법

현장에서 수거한 지문을 보관하고
있는 수억 개의 지문과 대조해서
같은 것을 찾아라!



어느 세월이 하나하나 비교하지?

CCTV 영상의 노이즈 제거하기

Noise reduction



Original Image



Modified Image



영상이 선명하지 않은데?

같은 수학 이론으로 해결 가능!

AlphaGo

인간의 직관에 해당하는 패턴인식 알고리즘 사용

→ 수읽기에서 경우의 수를 획기적으로 줄이는 데 성공 (우주의 원자보다 많은 가능성을 고려하는 대신 10만개 정도의 시나리오만 고려)

정책망: 착수를 결정하기 위해 탐색해야 하는 폭을 줄여서 확률이 높은 서너가지 착점만 고려

가치망: 현재의 형세가 누구에게 유리하지를 판단하기 위해 탐색해야 하는 깊이를 줄임.

정확한 값은 모든 경우의 수마다 게임 종료시까지 가봐야 해서 계산이 불가능

정책망과 가치망에서 영상 패턴을 인식하는 컨볼루션 신경망과 몬테카를로 탐색 사용

지도학습(supervised learning, learning by reading & studying)과

강화학습(reinforcement learning, leaning by doing)이 주효

알파고의 메시지

방대한 데이터 + 수학적 알고리즘



혁명적 변화

차원의 저주

Curse of Dimensionality

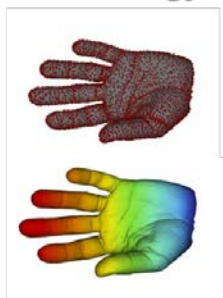
기계학습 알고리즘의 한계

- 4차 산업혁명은 가상세계와 실물세계의 연결이 특징
- 이러한 연결의 출발은 실물세계의 디지털화를 통한 빅데이터 확보이며, 실물세계 데이터의 수집 과정에서 사물인터넷의 역할 중요
- 이미지 데이터를 이용한 기계학습 등과는 달리, 실물세계의 디지털화 과정을 통해 확보한 빅데이터는 높은 차원을 갖는 경우가 많음
- 이 경우는 계산복잡도의 급격한 증대로 인해 기계학습이나 딥러닝 알고리즘의 단순 적용이 어려워짐
- 따라서 수학적 엄밀성을 갖춘 차원 축소 기법이 필요해지는데, 데이터 기반 최적화 기법을 위한 차원축소 방법 연구가 중요함

Mathematical Data-Analytics

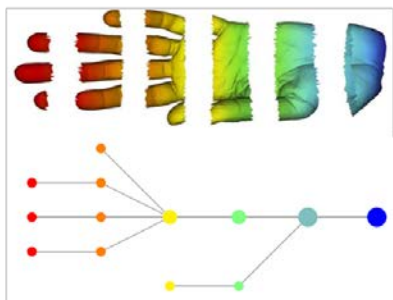
- 현대수학 이론의 도입으로 데이터의 본질적인 고유 성질을 수학적 불변량으로 표현 (Computational Topology)
- 통계적 또는 전산적 방식의 빅데이터 접근dmfh 찾아낼 수 없는 shape of data 정보를 이용하는 수학의 차별적인 빅데이터 접근법이며, 방대한 양의 고차원 데이터를 체계적으로 통제하고 다룰 수 있음

■ Computational Topology : **Simplicial Complexes,** Persistent Homology



A 3D object (hand) represented as a point cloud.

A filter value is applied to the point cloud and the object is now colored by the values of the filter function.



The data set is binned into overlapping groups.

Each bin is clustered and a simplicial complex is built.

- Pattern/Object Recognition

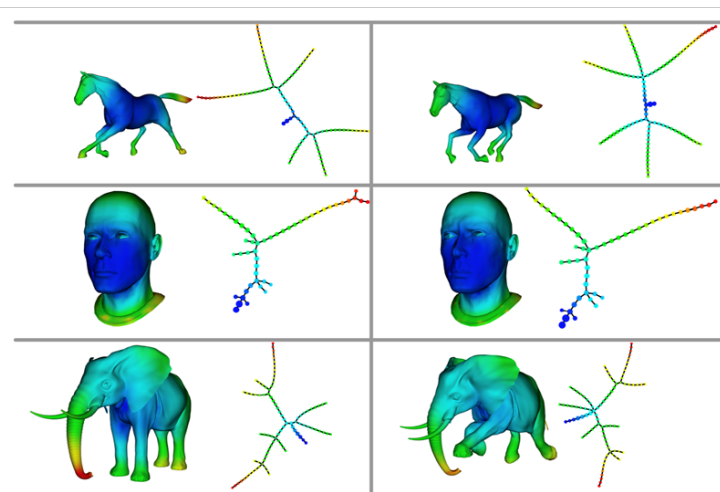


Figure 9: Refer to Section 5.3 for details. Each row of this image shows two poses of the same shape along with the Mapper result which is computed as described in Section 5.3. For each Mapper computation, we used 15 intervals in the range of the filter with a 50% overlap.

Mathematical Data-Analytics

- Mathematical can handle high-dimensional data, intrinsically. (Optimization)

- **Optimization: Nonnegative Matrix Factorization (NMF)**

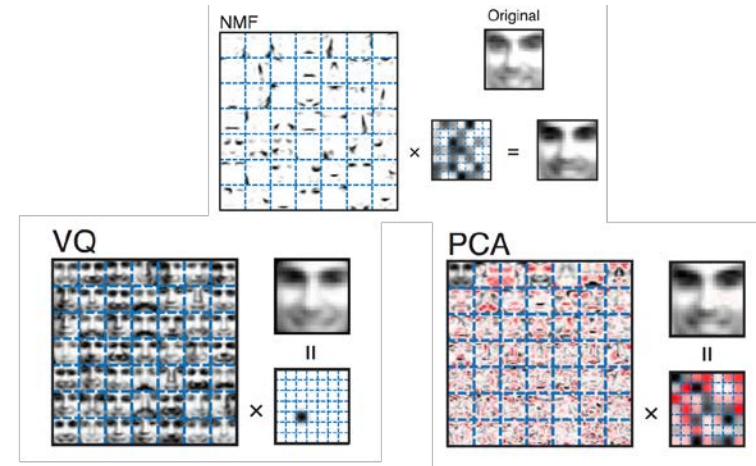
Given matrix $A \in \mathbb{R}^{m \times n}$, Find nonnegative matrices $W \in \mathbb{R}^{m \times k}$ and $H \in \mathbb{R}^{k \times n}$ satisfying

$$A \approx WH.$$

$$\min_{W \geq 0, H \geq 0} f(W, H) = \frac{1}{2} \|A - WH\|_F^2,$$

- Applications

- Text mining
- Image analysis and computer vision
- Social network
- Bioinformatics-microarray data analysis
- Acoustic signal processing
- Financial data
-

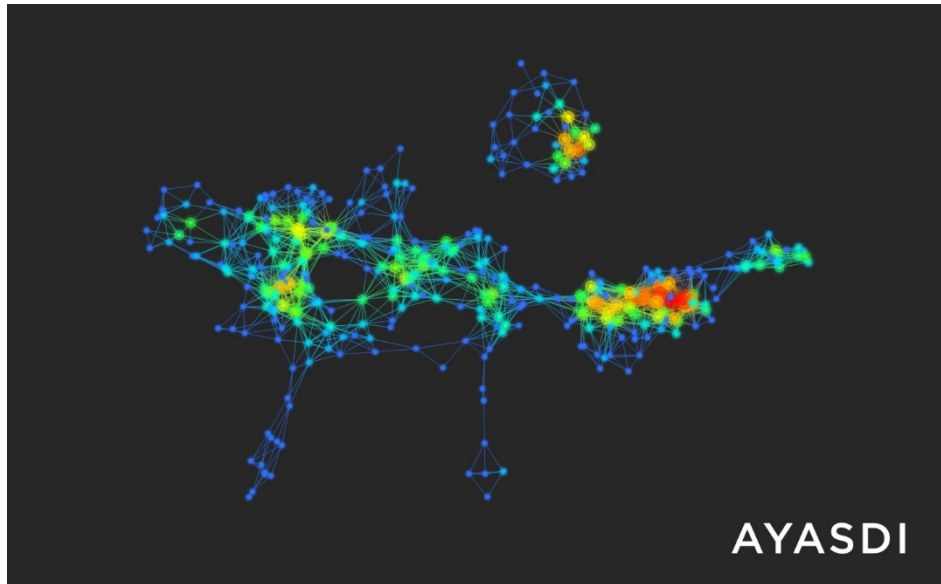


H. Kim and H. Park, Nonnegative Matrix Factorization Based on Alternating Non-negativity-constrained Least Squares and the Active Set Method. SIAM Journal on Matrix Analysis and Applications, 30(2):713-730, 2008.

Lee&Seung, Nature, 99

실리콘밸리의 TDA 스타트업

- **AYASDI** - 1,150억 원 이상의 투자 유치-



스탠포드 수학자 Gunnar Carlsson은

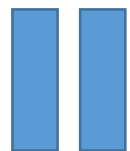
모양이나 형태를 연구하는 **위상수학**을 이용한
데이터분석기법 TDA(Topological Data
Analysis)를 창안

- ➔ 유방암 가능성 판단의 정확도 크게 증가
(Remember 안젤리나 졸리?)
- ➔ 당뇨병 Type 2 유무 및 6가지 방식의
상이한 처방 중에 무엇이 적절한지까지
자동으로 처방!!
- ➔ 기존에 숫자 위주의 데이터분석법으로는
해결할 수 없었던 의료와 금융 분야의
난제들을 해결

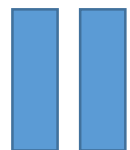
4차 산업혁명시대 공공 인프라

문제 해결 인프라 구축

산업수학



수학적 방식의 세상 문제 해결



문제해결 인프라

영화 산업의 예



Ron Fedkiw (1968년~)

아카데미 상을 두 번 받은 수학자



영화 ‘캐리비아의 해적’ 에서 태풍이 불어 물이 배로 넘치는 장면은 론 페드키의 소프트웨어를 사용해서 유체역학 방정식을 수치적으로 풀어 그려낸 것

교통



- 독일은 통일 직후 수도를 본에서 베를린으로 이전
- 동 베를린과 서 베를린이 통합된 수도 베를린의 대중교통 문제 심각
- 만차된 버스를 보내고 하염없이 다음 버스를 기다리는 시민의 긴 행렬
- 교통 혼잡으로 출퇴근 시간 크게 증가
- 통상적 접근: 버스를 증차하자. 버스노선을 늘리자.

- 베를린시는 대중교통 문제 해결방안 공모 실시
- 공모에서 선정된 연구팀(Martin Grötschel 교수)은 Integer programming 방식의 최적화 이론을 적용해서 교통 현황을 반영한 버스 노선의 대대적 변경 및 가변 배차시간표 도입
- 1,800대의 버스를 1,300대로 줄이고도 시민 대기시간은 획기적으로 줄이는데 성공

산업수학 점화프로그램

- 국내의 산업수학 기반이 미미한 상태에서 수학기계의 에너지를 세상 문제의 해결에 돌리는 방안 필요. 대학 교수들이 수학을 활용하여 기업이 현장에서 직면한 다양한 문제를 찾아서 해결하거나 유망 상품개발을 지원하는 프로젝트
- [기간 및 규모] 2015. 7. 1 ~ 2016. 12. 31 (총 18개월), 총 27억 원



▶ [현황]

- 국내 21개 대학
- 기업과 공동 프로젝트수행
- 산업수학 커리큘럼 개발
- 금융, 바이오, 빅데이터, 제조업, 딥 러닝등 다양한 분야

산업수학 관련 프로그램

- 산업수학센터(IMC) 선정: 2017년 2개의 센터를 대학에 설치하는 연구재단 프로그램이 진행 중. 2018년에도 선정 예정.
- [기간 및 규모] 총 5년, 연 10억원 규모. 확대 가능성 있음
- 산업수학 연구 공모 프로그램: 2017년 하반기에 총 50억원 규모의 연구비 지원 예정

산업수학혁신센터 (판교)



- 2016년 3월 개소하여 판교 테크노밸리의 스타트업캠퍼스 내 위치
- 산업현장의 문제를 수학적 방식으로 해결하는 전진기지의 역할
- 스타트업은 대기업에 비해 자체 R&D 인력이 부족하고 문제해결의 속도가 중요
- 센터 독자적 해결 또는 대전 연구소에서 문제해결포럼을 개최하여 빠른 문제 해결 지원
- (목요미팅) 스타트업을 초청하는 정례적 문제 발굴 활동

산업수학혁신센터 최근 문제 발굴/해결 사례



- 고리원자력발전소 핵연료 교체 문제
- 빅데이터 기반 인력매칭서비스
- 기계학습을 이용한 공작기계(CNC) 불량 탐지
- 가스회사 순찰차량 최적 순찰 경로
- 공압(pneumatic) 부품 최적 비축
- etc