

Big Healthcare Data Analytics

Challenges & Applications

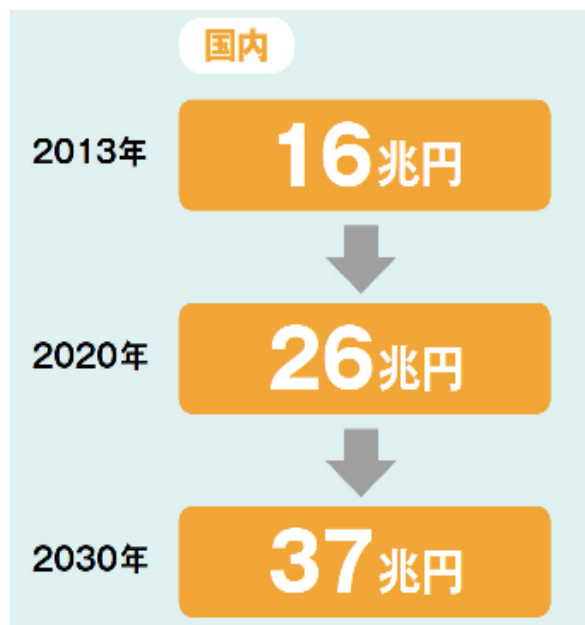
大阪大学サイバーメディアセンター
先進高性能計算機システムアーキテクチャ共同研究部門

Chonho Lee

@Cyber HPC Symposium, 2017

Healthcare (市場拡大)

市場規模



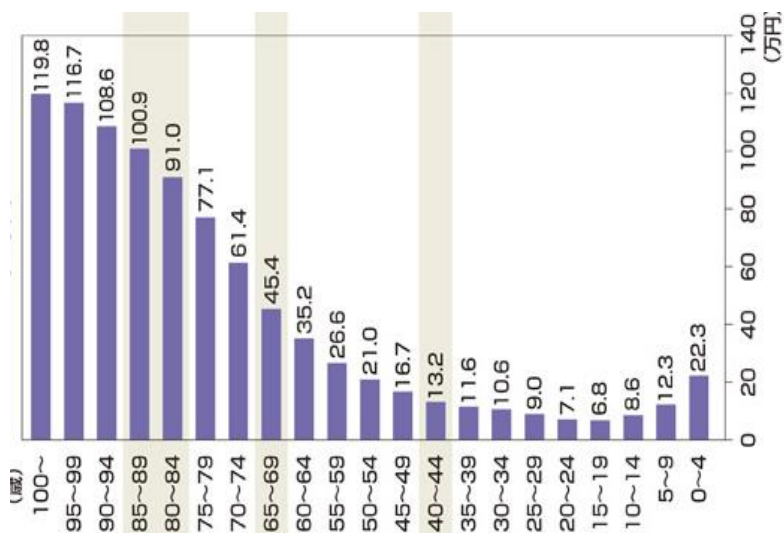
医療(ウェアラブル)機器市場



出典：日本再興戦略

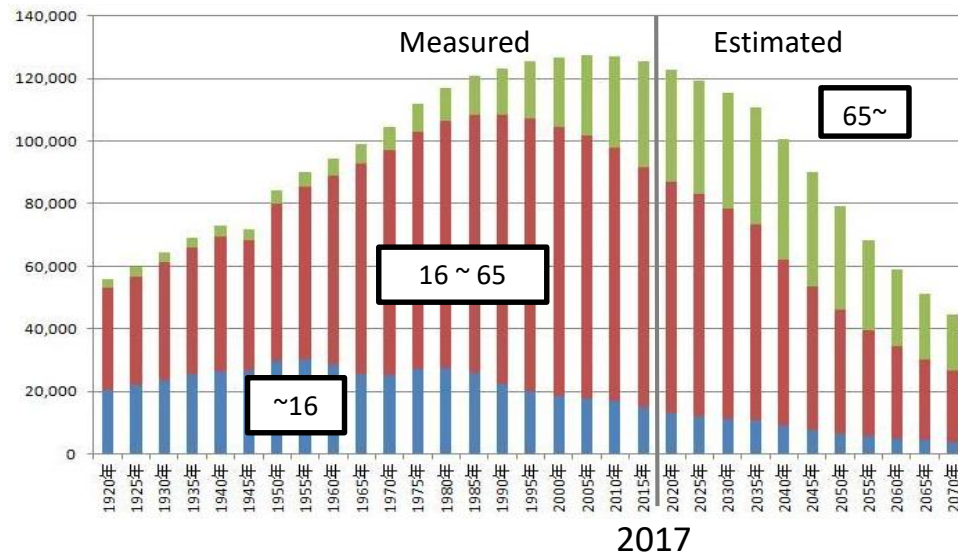
Healthcare (社会問題)

年間医療費



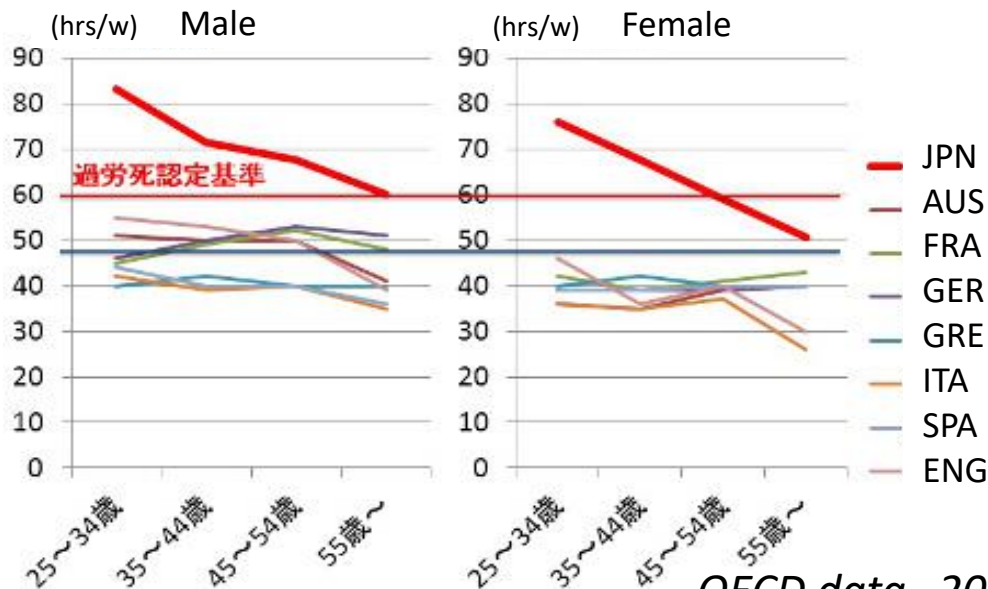
Ministry of Health, Labour and Welfare (2011)

人口推移

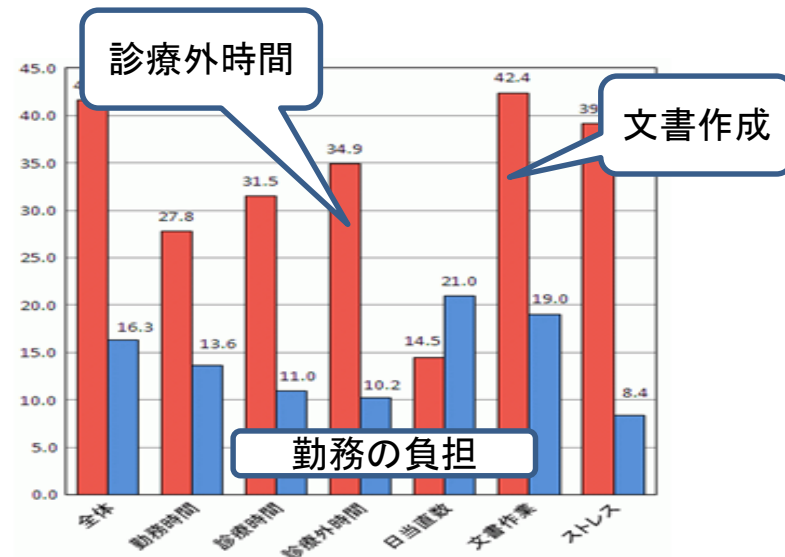


Ministry of Internal Affairs and Communications (2013)

Motivation



OECD data、2006



Dr-Union, Survey, 2013

- Most of the medical practices are completed by medical professionals **backed by their experiences**.
 - Clinical researches are conducted by researchers via **painstaking designed** and **costly experiments**.
- ➔ Can AI and HPC technologies support them ?



ビッグヘルスケアデータ



25億GB / 日（約9千億GB / 年）

90万PB / 年



ヘルスケアデータ

In 2012, worldwide digital healthcare data was estimated to be equal to 500 petabytes and is expected to reach 25,000 petabytes in 2020.

Hersh, W. (2011), Health-care hit or miss? *Nature*

知見を得る

本質を見抜く

価値を見出す

健康状態を知る

予防医療に役立てる

医療費削減につなげる

サイズの問題だけじゃない

ビッグデータの“4V”： Volume, Velocity, Variety, Veracity



複雑化

ヘルスケア データの種類

Electronic Healthcare Records (電子カルテ)

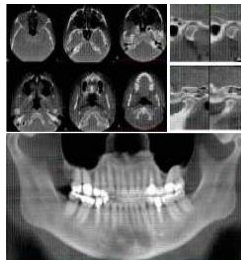


Structured Data

Demographics

Name, Age, Gender, Ethnic, Income, Education, etc.

検査結果、服薬記録、etc.



Unstructured Data

画像 : X-ray Image, CT scan, MRI

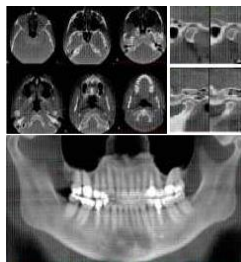
テキスト : 所見、治療計画書、経過観察、etc.

ヘルスケア データの種類

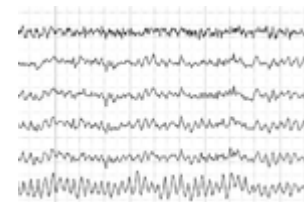
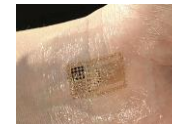
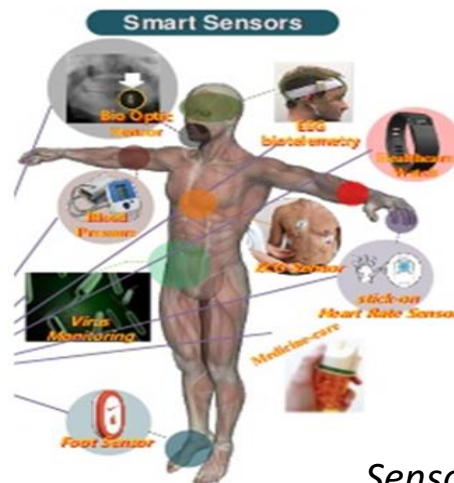
生体データ



Structured Data



Unstructured Data



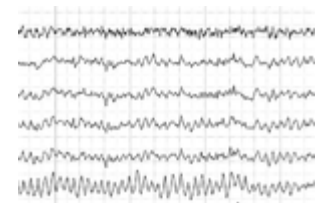
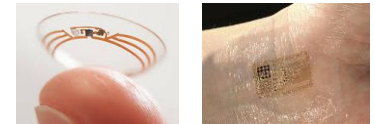
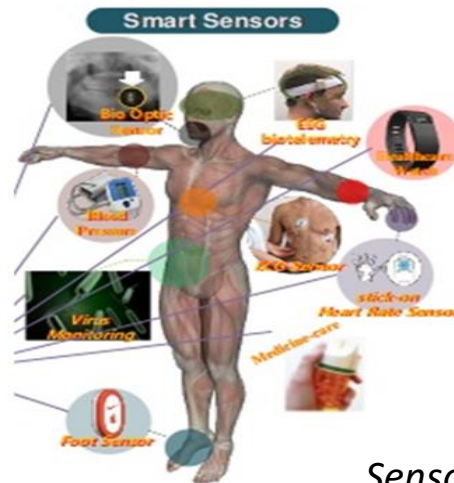
Sensor Data

加速度、ジャイロ
体温、心拍、血圧
ECG(心電)、EMG(筋電)、EEG(脳波)
涙の量、汗の成分、匂い、etc.

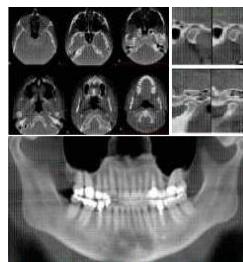
ヘルスケア データの種類



Structured Data



Sensor Data



Unstructured Data

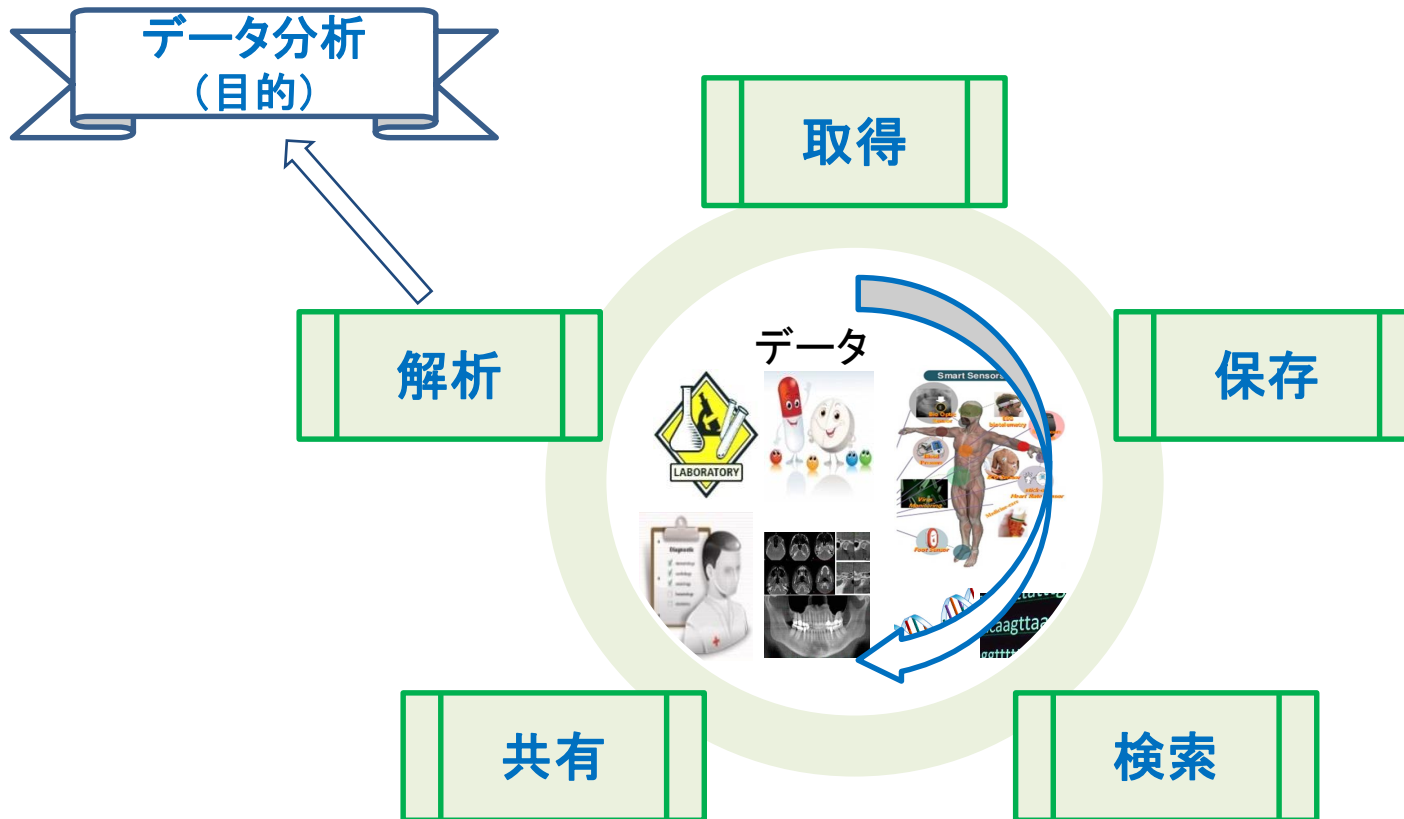


Genome Data

+ SNS data (Geological、Psychological information)

データ分析の 5 Do's

5 Do's: 効率よく、正確に



Challenges in Big Data Analytics

5 Do's: 効率よく、正確に

What & How: 何を、どのように

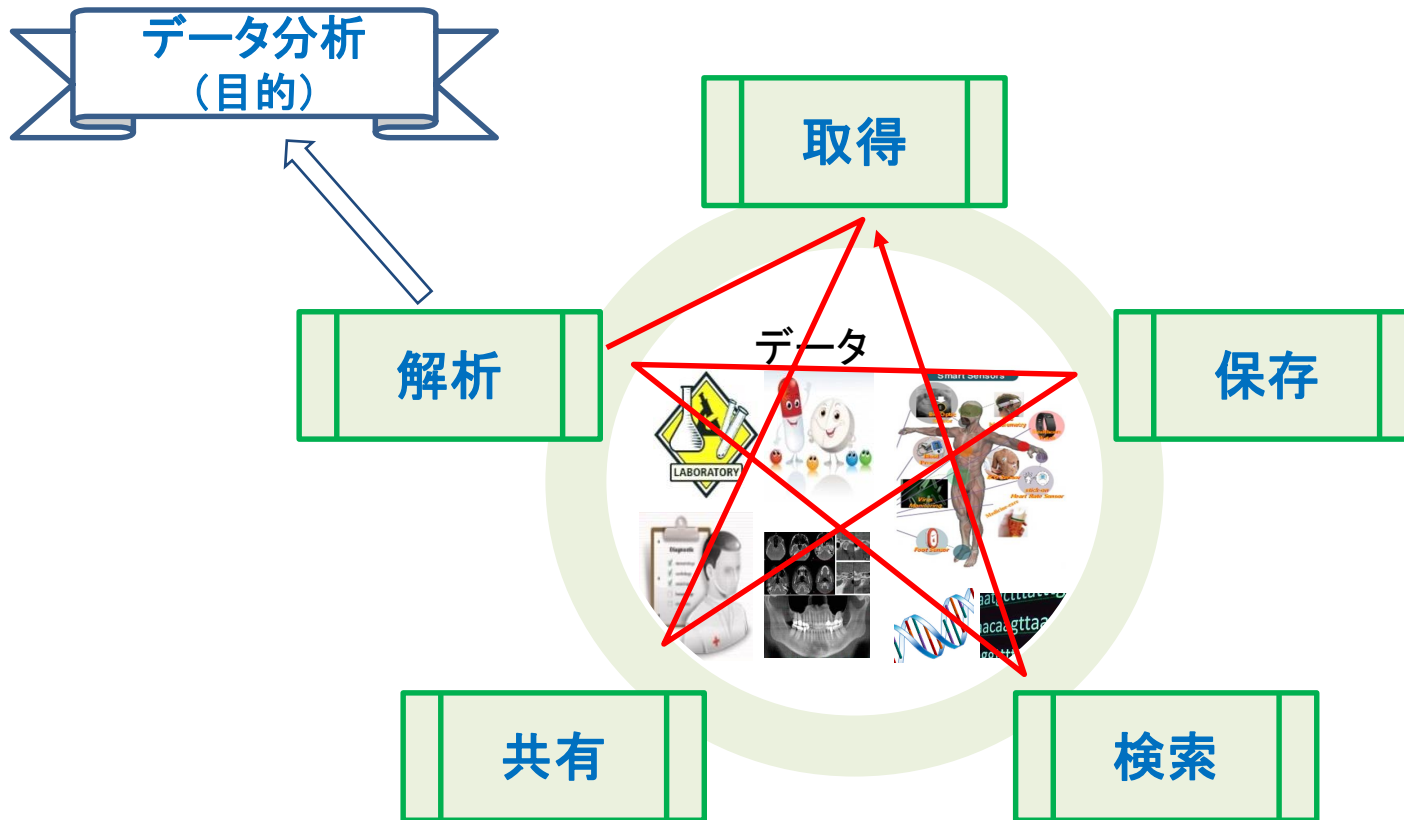
データと目的に合わせて色んなことを決めなければ、、、



Challenges in Big Data Analytics

5 Do's: 効率よく、正確に

What & How: 何を、どのように



実際は前処理に時間がかかる。行ったり来たり、、、

Challenges in Big Data Analytics

5 Do's: 効率よく、正確に What & How: 何を、どのように

Enterprise Data Analysis and Visualization: An Interview Study

Sean Kandel, Andreas Paepcke, Joseph M. Hellerstein, and Jeffrey Heer

I spend more than half of my time integrating, cleansing and transforming data without doing any actual analysis.

Most of the time I'm lucky if I get to do any analysis. Most of the time once you transform the data you just do an average... the insights can be scarily obvious.

Issues: High-dimension & Sparsity (1/3)

- EHR typically consists of hundreds to thousands of **medical features** (特徴) from multiple sources.

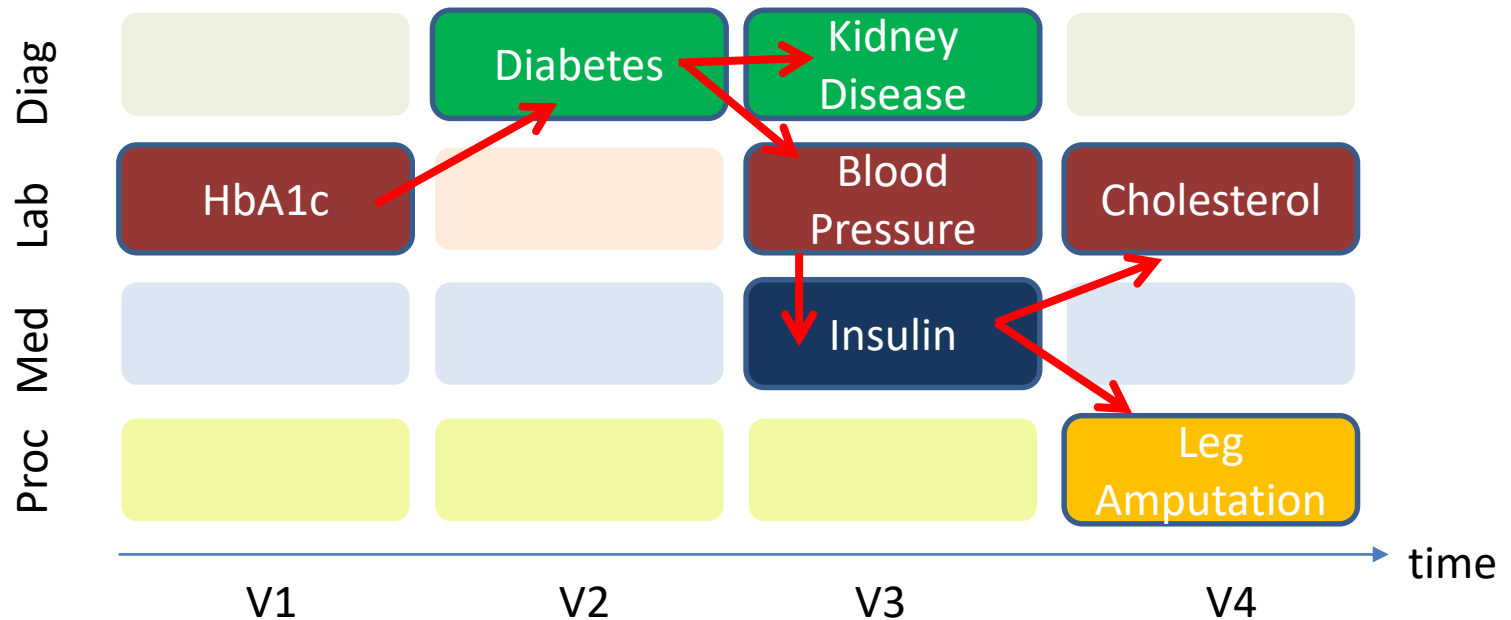
– 4 main category:

Diagnosis

Lab test

Medication

Procedure



Issues: High-dimension & Sparsity (2/3)

- EHR typically consists of hundreds to thousands of **medical features** (特徴) from multiple sources.

– 4 main category:

Diagnosis

Lab test

Medication

Procedure

- 例: National University Hospital in Singapore

– 約 10,000 患者の 1 年間のデータ

– 4143 特徴 (i.e., Diagnosis codes)

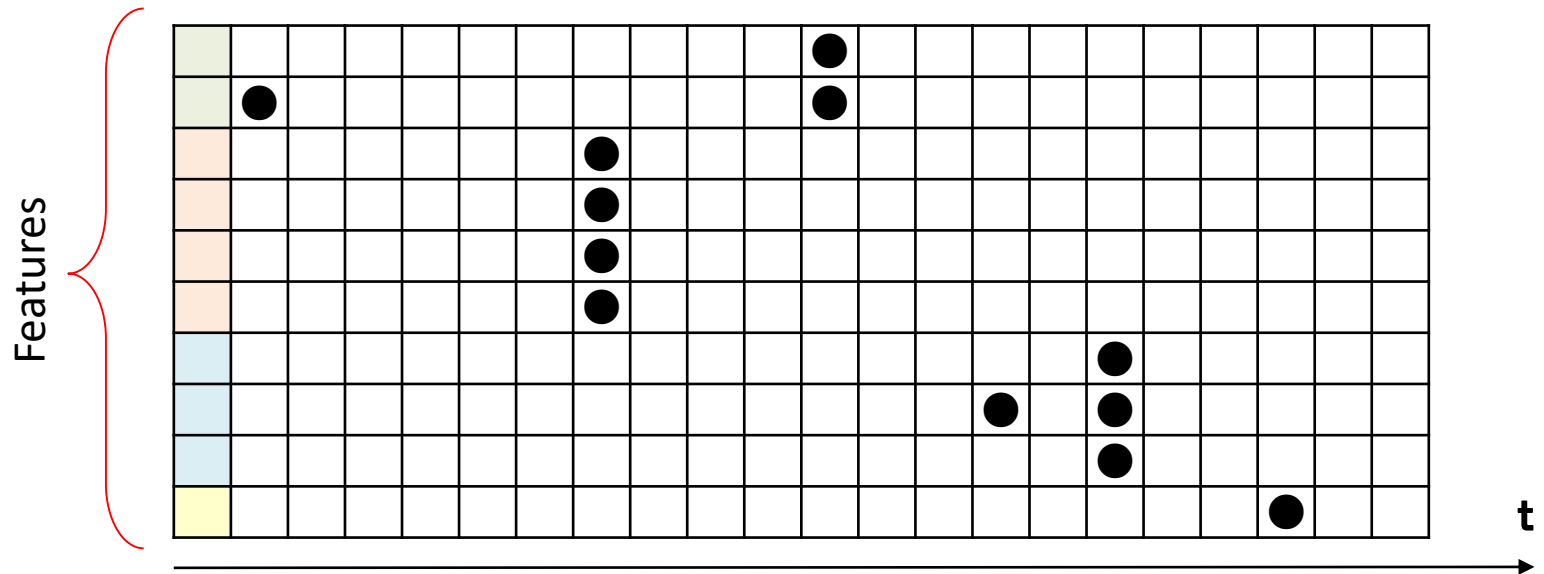
- 国際疾病分類 ICD-10 (International Statistical Classification of Diseases and Related Health Problems)

– about 80% of the patients have fewer than 10 codes

– about 70% of the patients have fewer than 4 visits to the hospital

Issues: High-dimension & Sparsity (3/3)

Illustration of high-dimensional and sparse dataset for a patient



of samples $\ll$ # of features

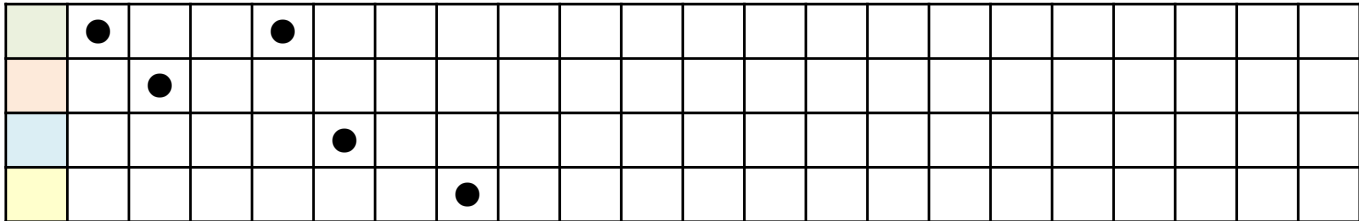
e.g., NUH datasetの、腎臓疾患を持つ糖尿病患者の1年間のデータ

- about 3000 patients v.s. 8000 features

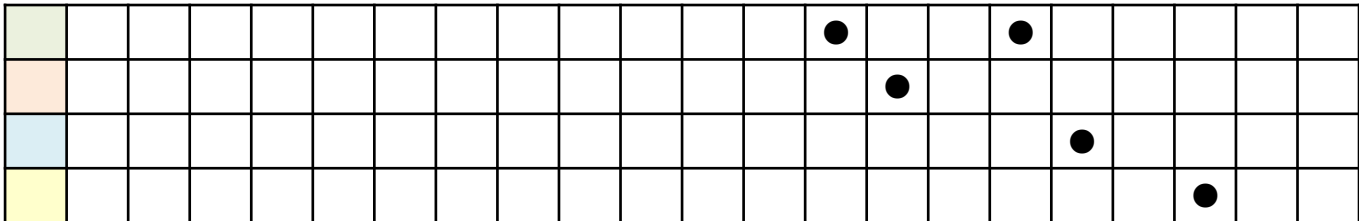
Issues: Irregularity

- 患者ごとに診療記録の time scale or granularity が違う

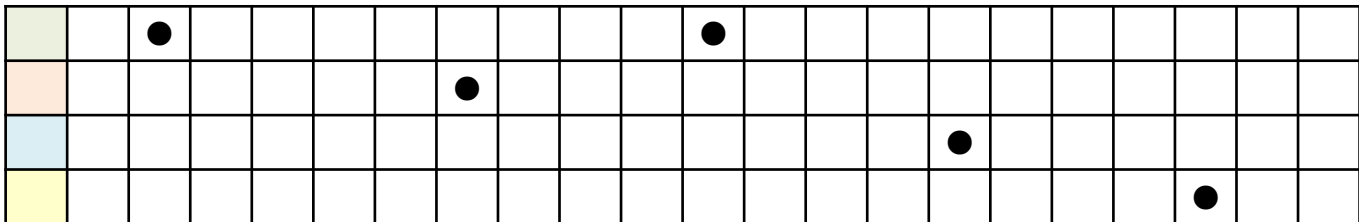
患者A



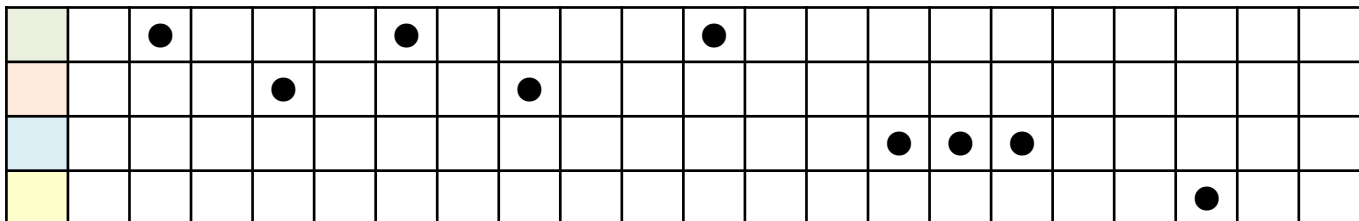
患者B



患者C



患者D



Issues: Missing data, Noise, Bias

- 機器精度、ヒューマンエラー、医師の主観による欠損値、ノイズ、偏り
 - 埋め合わせ・取り除き等の処理が増える。
 - Outlierが実はOutlierでない。
 - 「正解データ」が実は正解じゃない。
 - 患者は「陽性」に敏感
 - Positive Sampleが増えるが、Negative Sampleが少ない。
- Imputation（データ補完）による解決策
 - 統計的・確率的手法などがあるが、、、
 - 「ない」ものを「ある」として扱っていいのか？
 - 無視してもいいのか？
 - データが値なら Interpolation（補間）でいけそうだが、テキストデータはどう扱うの？

Framework for Big Healthcare Data Analytics

Data Analytics

Healthcare Services & Applications

Visualization



Complex Data Analytics Methods

Feature Selection

Analytic Model Selection

Predictive Analysis

Statistical Learning

Deep Learning

Correlation Learning

High Performance Computing Resources



Data Curation



Raw Data



Curated Data



Knowledge base



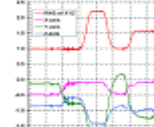
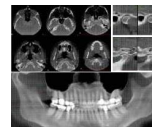
Medical Knowledge/Ontology



Structured Data



Unstructured Data



Sensor Data



Data Extraction

Data Cleansing

Data Annotation

Data Integration

Framework for Big Healthcare Data

Data Analytics

Healthcare Services & Applications

Visualization



Complex Data Analytics

Feature
Selection

Analytical
Model
Selection

Statistical
Learning

Deep Learning

Analysis

Correlation
Learning



Apache Singa

High Performance
Computing Resources



Data Curation



Raw
Data



Curated
Data



Knowledge
base



Medical



Structured Data



Unstructured Data



Sensor Data



CROWDSOURCING DATA ANALYTICS SYSTEM

Data
Extraction

Data
Cleansing

Data
Annotation

Data
Integration

Example of Data Extraction・Cleaning

Case Note (text) → Data Extraction → Data Cleansing

例

Pt given milo at triage at 0750hrs. BSL 3.7 mmol upon triage.

PMH

58 yo/Chinese/M

Ambulant with quadstick since Feb 2012 after CVA

1)DM

- Last Hba1c 6.2% (31/5/12)
- On metfomin 500mg BD and glipizide 10mg BD
- Cx by

⋮

R1: Error correction (誤字)

R2: Ambiguity removal (曖昧さ)

R3: Relation discovery/construction (関連)

R4: Filling missing values (欠損値)

Pt 何？

0750 時間？

BSL 何？

3.7mmol 大丈夫なの？

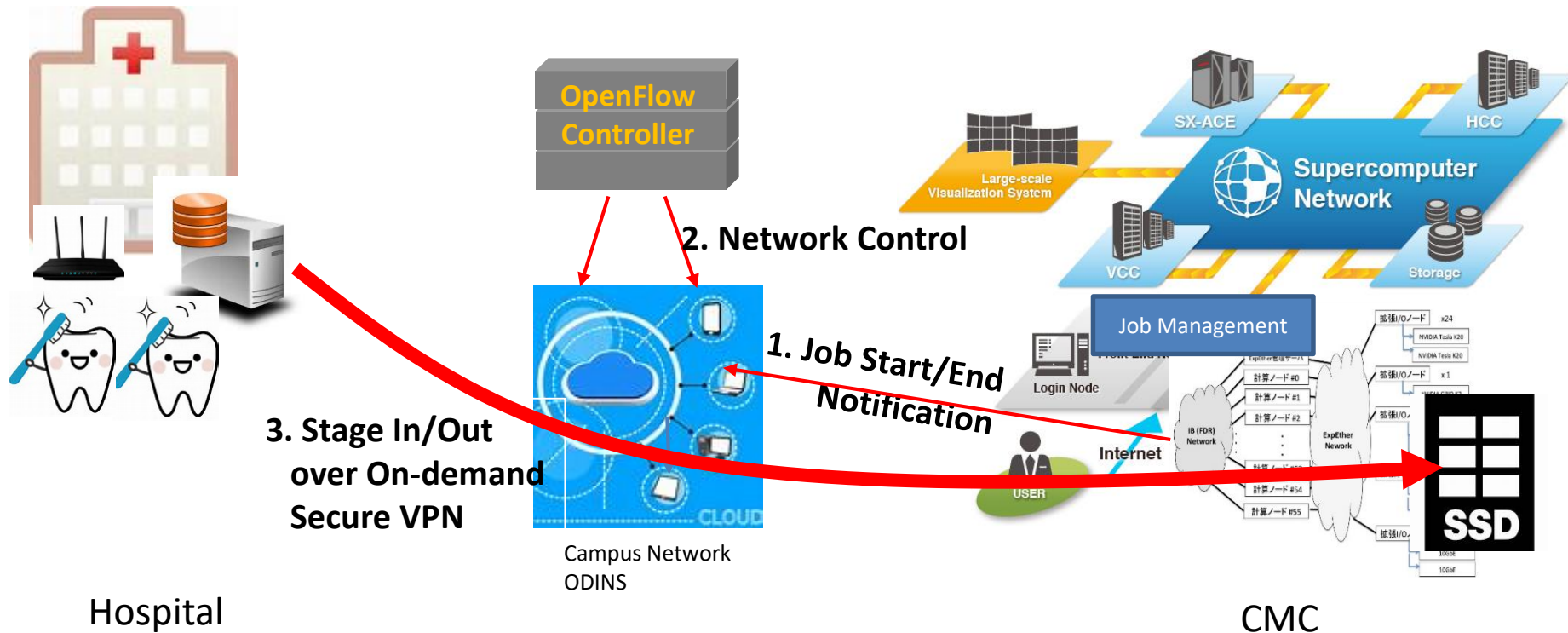
CVA 文章から推測しなければ

Metfomin -> Metformin

Secure VPN for Healthcare Data Analytics

On-demand Secure Stage-in/out

- **Connect CMC-Group network and locate/transfer data to compute nodes when only needed.**
- Disconnect network and delete data right after**



Example Application: 矯正歯科治療

Dataset

Facial photo



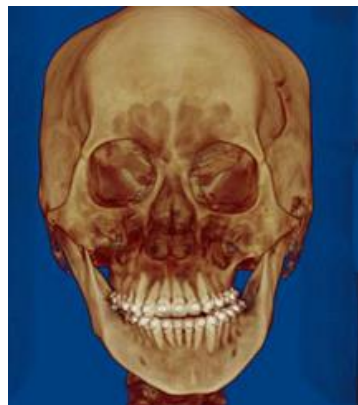
Oral photo



X-ray: Cephalogram



CT (DICOMデータ)



OBJデータ



顔画像所見

正面観：非対称である。

オトガイの右方への偏位を認める。

鼻尖の左方への偏位を認める。

笑顔表出時には、上顎前歯の歯冠全体と下顎前歯の歯冠の約1/3の露出を認める。

安静時には、上顎前歯の歯冠の約6mmの露出を認める。

側面観：convex type profileである。

口唇閉鎖時および安静時にオトガイの緊張を認める。

E line：上赤唇はE lineより約8.0mm前方に位置している。

下赤唇はE lineより約11.5mm前方に位置している。

鼻唇角：鼻唇角は大きい。Nasolabial angle: 105°

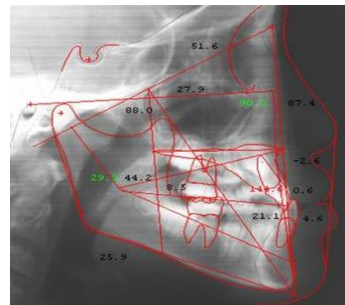
顔面の垂直的バランス：前中顔面の長さに対して、前下顔面の長さは標準的である。

鼻下点から口裂までの長さに対して、口裂からオトガイまでの長さは標準より短い。

gl-sn:sn-me=1:1, sn-sto:sto-me=1:1.67

Demographics: 名前、年齢、性別など

Textデータ: 治療前所見、治療計画、経過観察など



主訴 (Chief complaint)

口唇腫
 唇瘡 (lipogumma)
 angle chancr
 (Shenka class 2)

① 唇の炎症の状況 (Inflammation)
 ・上唇の唇瘡の存在は、やや軽度 (舌側) 緑色している。約 5 度
 ・下唇の唇瘡の存在は、やや軽度 (舌側) 緑色している。約 8 度
 ・上唇の唇瘡の炎症の程度は、左右ともに下唇の唇瘡が軽度にすぎた状態では不安定である。

② 歯の状態 (Dental)

・上顎前歯の歯冠は、やや軽度 (舌側) 緑色している。約 5 度
 ・下顎前歯の歯冠は、やや軽度 (舌側) 緑色している。約 8 度
 ・上唇の唇瘡の炎症の程度は、左右ともに下唇の唇瘡が軽度にすぎた状態では不安定である。

③ 歯肉の炎症の状況 (Gingivitis)

・上唇前歯の歯冠は、やや軽度 (舌側) 緑色している。約 5 度
 ・下顎前歯の歯冠は、やや軽度 (舌側) 緑色している。約 8 度
 ・上唇の唇瘡の炎症の程度は、左右ともに下唇の唇瘡が軽度にすぎた状態では不安定である。

④ 歯肉の炎症の状況 (Gingivitis)

・上唇前歯の歯冠は、やや軽度 (舌側) 緑色している。約 5 度
 ・下顎前歯の歯冠は、やや軽度 (舌側) 緑色している。約 8 度
 ・上唇の唇瘡の炎症の程度は、左右ともに下唇の唇瘡が軽度にすぎた状態では不安定である。

治療計画 (Treatment planning)

① 唇の炎症の状況 (Inflammation)
 ・上唇の唇瘡の炎症の程度は、左右ともに下唇の唇瘡が軽度にすぎた状態では不安定である。

② 歯の状態 (Dental)

・上唇前歯の歯冠は、やや軽度 (舌側) 緑色している。約 5 度
 ・下顎前歯の歯冠は、やや軽度 (舌側) 緑色している。約 8 度
 ・上唇の唇瘡の炎症の程度は、左右ともに下唇の唇瘡が軽度にすぎた状態では不安定である。

③ 歯肉の炎症の状況 (Gingivitis)

・上唇前歯の歯冠は、やや軽度 (舌側) 緑色している。約 5 度
 ・下顎前歯の歯冠は、やや軽度 (舌側) 緑色している。約 8 度
 ・上唇の唇瘡の炎症の程度は、左右ともに下唇の唇瘡が軽度にすぎた状態では不安定である。

④ 歯肉の炎症の状況 (Gingivitis)

・上唇前歯の歯冠は、やや軽度 (舌側) 緑色している。約 5 度
 ・下顎前歯の歯冠は、やや軽度 (舌側) 緑色している。約 8 度
 ・上唇の唇瘡の炎症の程度は、左右ともに下唇の唇瘡が軽度にすぎた状態では不安定である。

⑤ 歯肉の炎症の状況 (Gingivitis)

・上唇前歯の歯冠は、やや軽度 (舌側) 緑色している。約 5 度
 ・下顎前歯の歯冠は、やや軽度 (舌側) 緑色している。約 8 度
 ・上唇の唇瘡の炎症の程度は、左右ともに下唇の唇瘡が軽度にすぎた状態では不安定である。

⑥ 歯肉の炎症の状況 (Gingivitis)

・上唇前歯の歯冠は、やや軽度 (舌側) 緑色している。約 5 度
 ・下顎前歯の歯冠は、やや軽度 (舌側) 緑色している。約 8 度
 ・上唇の唇瘡の炎症の程度は、左右ともに下唇の唇瘡が軽度にすぎた状態では不安定である。

医師の作業負担軽減
診断の客観化
Informed Consentのための材料

- 重症度判定
- 所見生成
- 特徴点抽出、診断
- 治療計画提案
- 治療後予測

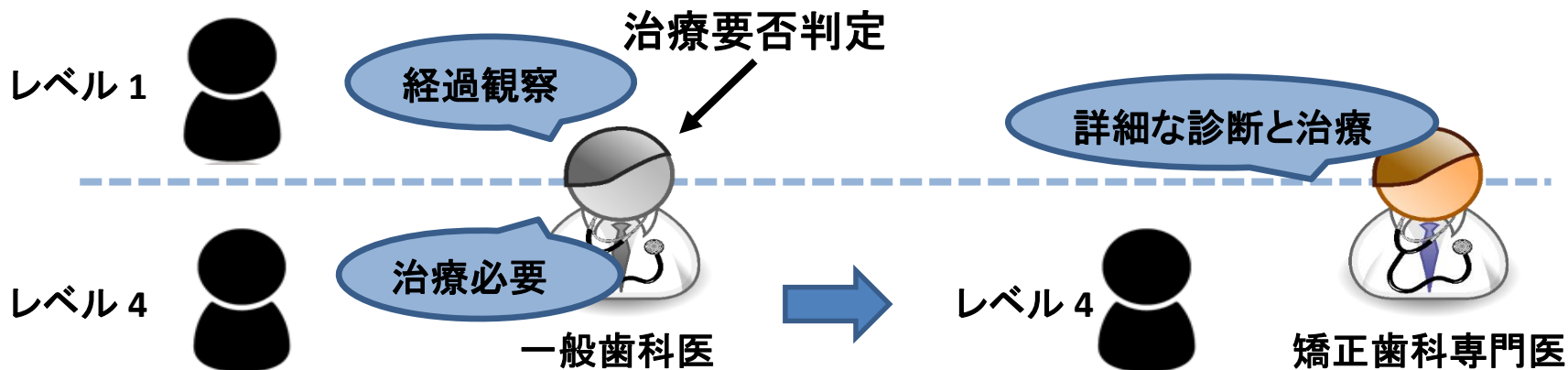
Example Applications: Severity Measurement

Index of Orthodontic Treatment Priority (IOTN)

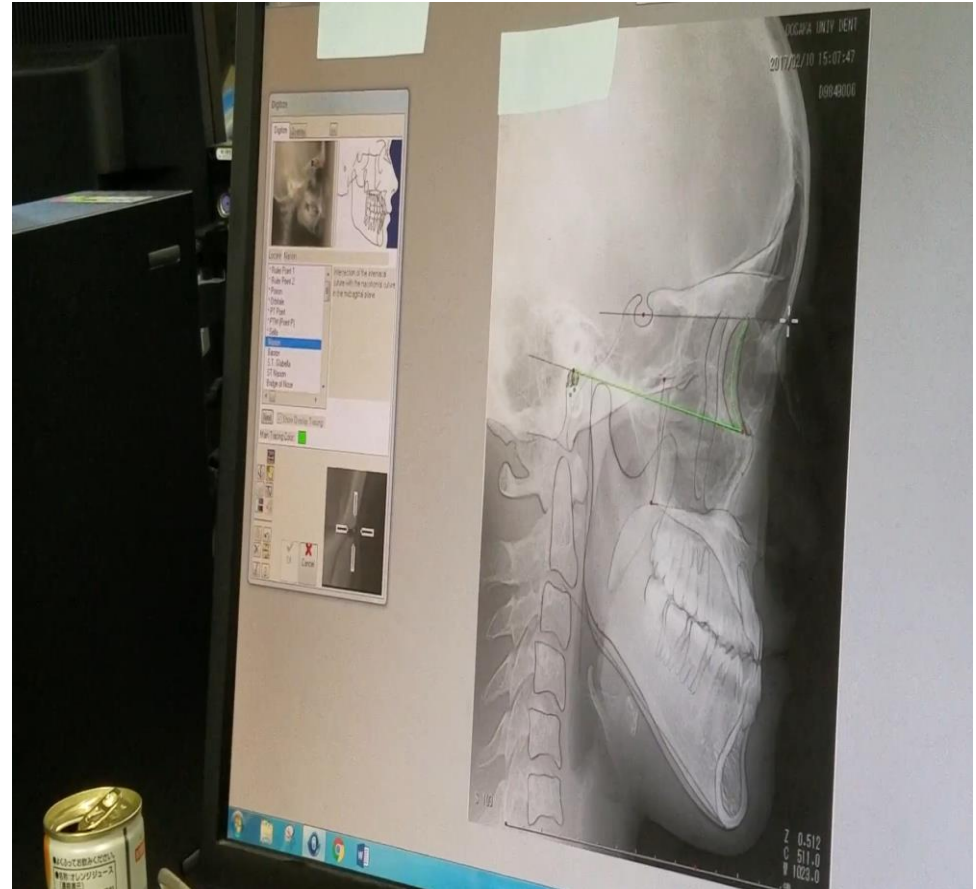
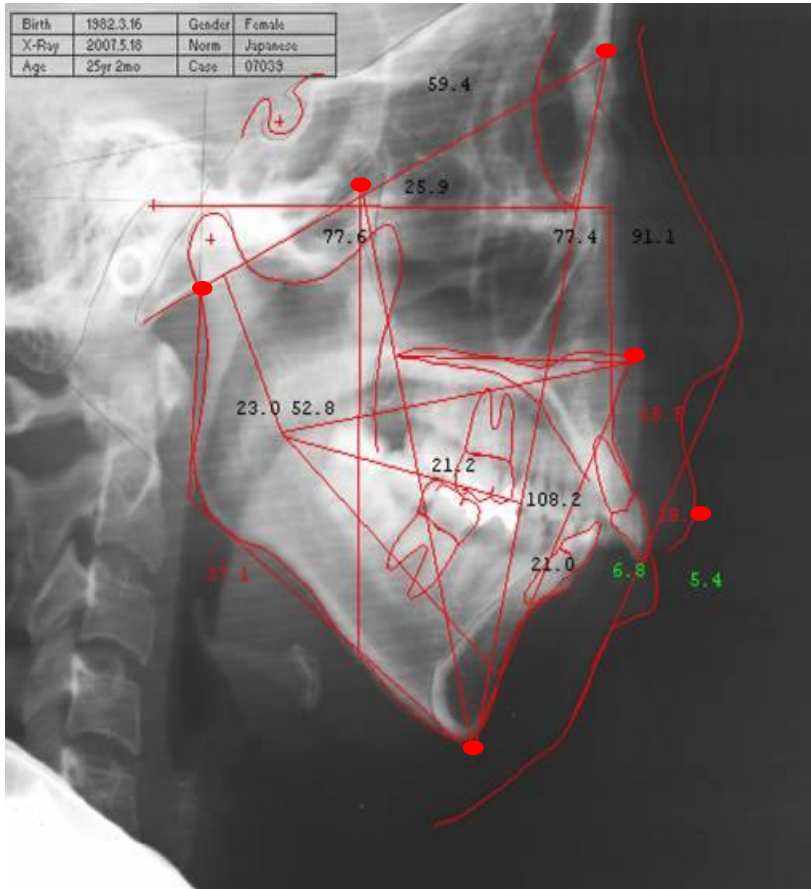
- malocclusion (不正咬合)や jaw abnormality (顎変形症)の度合い
- 治療介入の必要性を計る5段階評価基準



- 一般歯科医でレベル4,5と判定されれば矯正歯科専門医へ紹介



Morphological Landmarking



所見・治療計画書作成をサポート



画像出典:高田 健治 メデジットコー
ポレーション『高田の歯科矯正の学
び方 わかる理論・治す技術』, 2010.

現症

1. 全身所見
特記事項なし。

2. 顔画像所見
正面観：左右非対称である。
左側に比べて右側の目の位置が高い。
オトガイの偏位は認めない。
咬合平面の傾斜は認めない。
Smile時には、上顎前歯の歯冠2/3が露出している。
側面観：straight type profileである。
上顔面と下顔面の比率は下顔面が小さく、鼻下点から口裂と口裂からオトガイまでの比率は標準的である。
gl-sn:sn-me=1:0.8, sn-sto:sto-me=1:2.0
E line: 上唇はE lineより約1.0mm後方に位置している。
下唇はE lineより約1.0mm後方に位置している。
鼻唇角：鼻唇角は標準的である。
nasolabial angle: 105.0°

3. エックス線画像所見
(1) セファロ画像分析所見
(i) 正面位:
骨格性: 上顎骨の正中は顔の正中とほぼ一致している。
下顎骨の正中は顔の正中とほぼ一致している。
歯性: 上顎歯列の正中は顔の正中に対して0.5mm左側に偏位している。
下顎歯列の正中は顔の正中に対して1.5mm左側に偏位している。
咬合平面の傾斜は認めない。
Lo-Lo' /UR6: 79.5mm Lo-Lo' /UL6: 79.5mm (左右差約0.0mm)

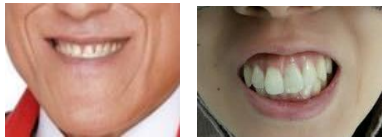
(ii) 側面位:
骨格性: 骨格性1級
ANB: 4.8° (0.8s.d.)
SNの長さは成人の女子の標準範囲内である。
SN: 70.1 mm (0.6s.d.)
上顎骨: 上顎骨の位置は標準範囲内である。前後径は大きい。
SNA: 81.5° (0.2s.d.), Ptm-A/PP: 50.7mm (1.0s.d.), Ptm-Ans/PP: 55.2mm (1.0s.d.)
下顎骨: 下顎骨の位置は標準範囲内である。下顎枝高、下顎骨体長と下顎実効長は標準的である。
SNB: 76.7° (-0.3s.d.), Go-Me: 73.3 mm (0.5s.d.), Ar-Me: 104.9 mm (-0.3s.d.)
Ar-Go: 46.6 mm (-0.2s.d.)
下顎下縁平面はlow mandibular plane angleを示す。下顎角は小さい。
FH-MP: 27.3° (-0.9s.d.), SN-MP: 32.3° (-1.0s.d.), GoA: 116.3° (-1.1s.d.)
前顔面高・中顔面高は標準的である。下顔面高は小さい。
N-Me: 122.1mm (-0.7s.d.), N/PP: 56.8mm (0.3s.d.), Me/PP: 64.9mm (-1.0s.d.)
歯性: U1は口蓋側傾斜している。
U1-FH: 85.0° (-3.3s.d.) U1-SN: 80.0° (-3.0s.d.)
L1の歯軸は標準的である。
L1-MP: 91.6° (-0.3s.d.) L1-FH: 61.1° (0.6s.d.)
U1, L1, L6の位置は標準的である。U6は低位である。
U1/PP: 29.2mm (-0.8s.d.) U6/PP: 20.9mm (-1.8s.d.) L1/MP: 41.7mm (-0.9s.d.) L6/MP: 34.5mm (0.7s.d.)
IIAは大きい。
IIA: 156.1° (3.1s.d.)
Over jetは標準的である。Overbiteは標準範囲内であるが、やや大きい傾向にある。
OJ: 3.2mm (0.1s.d.), OB: 5.1mm (0.9s.d.)

(2) パノラマおよび咬合法エックス線画像所見
38は水平埋伏している。その他の第二大臼歯までの全ての永久歯を認める。
下顎前歯部に歯周疾患による歯槽骨の吸収を認める。

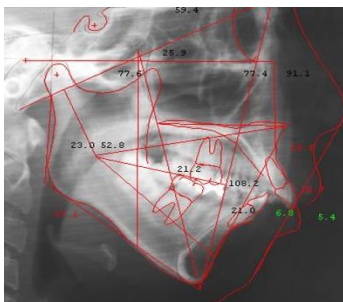
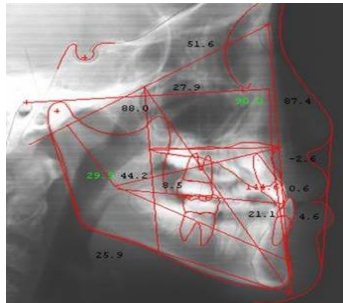
Input Data



⋮



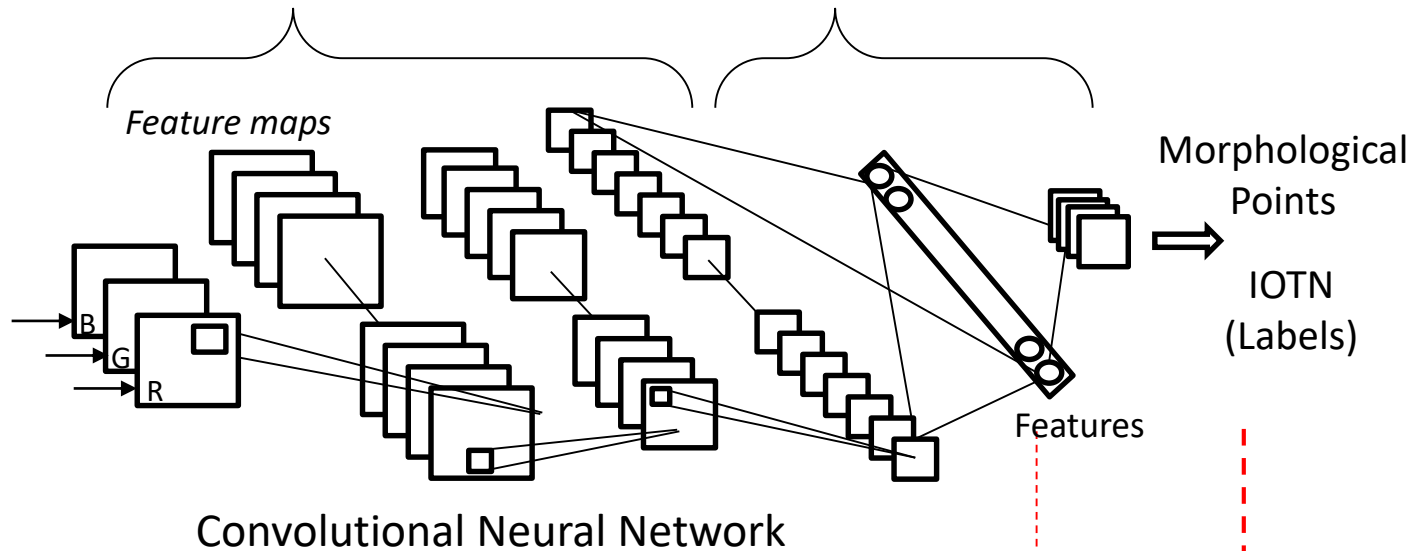
⋮



⋮

Convolution & Pooling operations

Multi-layer Perceptron



Output

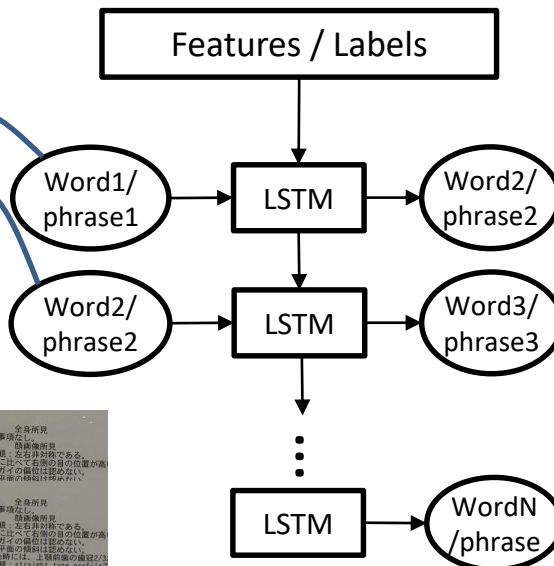
Lookup Table
Word
Embedding



*Knowledge base
(text documents)*

現在 全歯列
特記事項なし
正面観 左右非対称である。
左側の上顎の歯の位置が右
側の上顎の歯の位置と異なる。
現在 全歯列
特記事項なし
正面観 左右非対称である。
左側の上顎の歯の位置が右
側の上顎の歯の位置と異なる。
現在 全歯列
特記事項なし
正面観 左右非対称である。
左側の上顎の歯の位置が右
側の上顎の歯の位置と異なる。

Text Generating Recurrent Network



Thank You for listening!

Cybermedia Center & Graduate School of Dentistry in Osaka University

