

HadoopカンファレンスJapan 2013 Winter 2013年1月21日

～医療情報分野でHadoopをどう使う？～
最新事情と効果的な活用のためのアプローチ

水谷 晃三 澤 智博
帝京大学医療情報システム研究センター

Copyright (C) 2013 Kozo Mizutani, Teikyo University.

0. 本講演について

- ◆ 自己紹介
 - 元プログラマ
 - 専門
 - 医療情報システム分野, 教育工学分野
 - 現在の主な仕事
 - 病院情報システムの企画, 開発, 運用
 - 技術等の調査, 医療情報システムに関わる研究
 - 理工学部での教務 など

Copyright (C) 2013 Kozo Mizutani, Teikyo University.

0. 本講演について

- ◆ 本講演の目次
 - 1. 医療分野におけるHadoopの最新事情
 - 2. 効率的なデータ活用に向けて
 - 帝京大学医学部附属病院での取り組み
 - 3. 生体モニタデータへの適用事例
 - Hadoopを用いた類似性分析

Copyright (C) 2013 Kozo Mizutani, Teikyo University.

1. 医療分野における最新事情

- ◆ 医学分野の研究はここから
 - PubMed
 - <http://www.ncbi.nlm.nih.gov/pubmed>
 - 米国国立生物工学情報センター(NCBI)が運営する医学・生物学分野の学術文献検索
 - 世界の主要な約5,600雑誌, 約2200万件の文献

Copyright (C) 2013 Kozo Mizutani, Teikyo University.

1. 医療分野における最新事情

- ◆ PubMedによる検索サマリ
 - 「Hadoop」で検索
 - 25件 (2009年～)
 - 「MapReduce」で検索
 - 31件 (2008年～)

※2013年1月16日現在

Copyright (C) 2013 Kozo Mizutani, Teikyo University.

1. 医療分野における最新事情

- ◆ PubMedによる検索サマリ
 - 「Hadoop」25件の内訳
 - Bioinformatics 分野：19件
 - Medical Informatics 分野：4件
 - うち、医療画像関係が3件
 - その他, 総説など：2件

Copyright (C) 2013 Kozo Mizutani, Teikyo University.

1. 医療分野における最新事情

- ◆ 適用例が多いBioinformatics分野
 - Bioinformatics=生物情報学
 - データの蓄積、コンピュータによる分析が必須となる分野
 - DNAやタンパク質などの構成要素の配列
⇒膨大なテキストベースのデータ

Copyright (C) 2013 Kozo Mizutani, Teikyo University

1. 医療分野における最新事情

- ◆ 適用例が多いBioinformatics分野
 - DNAデータの例

```
ORIGIN
1 atggagagaa taaaagaact gagagatcta atgtgcgagt cccgcactcg cgagatactc
61 actaagacca ctgtggacca tatggccata atcaaaaagt acacatcagg aaggcaagag
121 aagaaccgcc cactcagaat gaatgggatg atggcaatga gataccaat tacagcagag
181 aagagaataa tggacatgat tccagagagg aatgaacaag gacaaacct ctggagcaaa
241 acaaacgatg ctggatcaga ccgagtgatg gtatcacctc tggccgtaac atggtggaat
301 aggaatggcc caacaacaag tacagttcat taccctaagg tatataaac ttatttcgaa
361 aaggtcgaaa ggttgaaca tggatcacttc ggccctgctc acttcagaaa tcaagttaaa
421 ataaagagga gagttgatc aaacctggc catgcagatc tcagtgccaa ggagcaacag
481 gatgtgatta tggaaagtgt ttcccaaat gaagtggggg caagaact gacatcagag
      :
      :
      :
```

※DBJ: [ACCESSION No.FJ966079](http://www.ncbi.nlm.nih.gov/nucl/ACCESSION.No.FJ966079) より

Copyright (C) 2013 Kozo Mizutani, Teikyo University

1. 医療分野における最新事情

- ◆ 適用例が多いBioinformatics分野
 - (個人的な見解ですが)
 - DNAデータベースのデータ形式が大きく変わることはないだろう
 - Hadoopのデータローカリティの有効性が得やすいデータの1つといえる

Copyright (C) 2013 Kozo Mizutani, Teikyo University

1. 医療分野における最新事情

- ◆ 医療情報の利活用の重要性
 - 臨床研究の立場など、医療情報の利活用の重要性は重ねて強調されてきた。
 - 近年の情報技術の発展により、医療情報の利活用が本格的になりつつある。
 - Bioinformaticsの発展と共に、医療情報の利活用の重要性が増している。

Copyright (C) 2013 Kozo Mizutani, Teikyo University

1. 医療分野における最新事情

- ◆ しかし実際は・・・
 - Hadoopに関わる文献は4件ほど
 - しかも、うち3件は医療画像に関するもの
 - 電子カルテの普及が遅れているという事情もあります・・・
 - 背景には医療情報システムに関わる(慢性的な)課題も。

Copyright (C) 2013 Kozo Mizutani, Teikyo University

1. 医療分野における最新事情

- ◆ 医療情報分野におけるHadoop



Hiromasa H., et al. A user-friendly tool to transform large scale administrative data into wide table format using a mapreduce program with a pig latin based script, BMC Medical Informatics and Decision Making 2012, 12:151より

Copyright (C) 2013 Kozo Mizutani, Teikyo University

1. 医療分野における最新事情

- ◆ 医療情報分野におけるHadoop
 - 日本医療情報学会他, 国内での取り組み
 - 石井, 他: 分散ファイルシステムを活用した多施設統合型臨床研究ストレージの設計, 2012
 - 吉田, 他: Hadoopを用いた電子カルテの統計分析とタイムライン検索, 2012
 - 演者らによるHadoopによる生体モニタの類似性分析, 2011~2013

Copyright (C) 2013 Kozo Mizutani, Teikyo University

1. 医療分野における最新事情

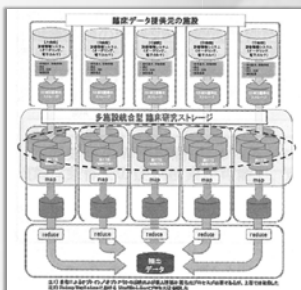
- ◆ 医療情報分野におけるHadoop
 - 日本医療情報学会他, 国内での取り組み



石井雅通, 大江和彦: 分散ファイルシステムを活用した多施設統合型臨床研究ストレージの設計, PB-1-2-3, 2012 より

Copyright (C) 2013 Kozo Mizutani, Teikyo University

1. 医療分野における最新事情



石井雅通, 大江和彦: 分散ファイルシステムを活用した多施設統合型臨床研究ストレージの設計, 第16回日本医療情報学会春季学術大会 PB-1-2-3, 2012 より

Copyright (C) 2013 Kozo Mizutani, Teikyo University

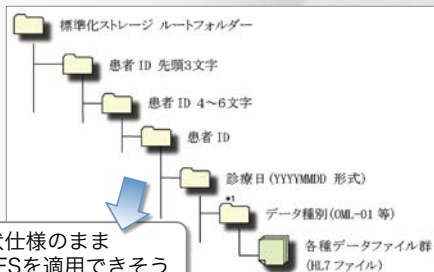
1. 医療分野における最新事情

- ◆ SS-MIXについて
 - 厚生労働省: 電子的診療情報交換推進事業
 - SS-MIX: Standardized Structured Medical record Information eXchange
 - さまざまなインフラから配信される情報を蓄積するとともに標準的な診療情報提供書が編集できる「標準化ストレージ」という概念に着目し、すべての医療機関を対象とした医療情報の交換・共有による医療の質の向上を目的とした事業

SS-MIX普及推進コンソーシアム
<http://www.hci-bc.com/ss-mix/ssmix/index.html#ssmix>

Copyright (C) 2013 Kozo Mizutani, Teikyo University

1. 医療分野における最新事情



現状仕様のままHDFSを適用できそう

SS-MIX2 標準化ストレージ構成の説明と構築ガイドライン Ver.1.0 rev1, P.5より引用 (一部抜粋)

Copyright (C) 2013 Kozo Mizutani, Teikyo University

1. 医療分野における最新事情

- ◆ 医療情報分野におけるHadoop
 - その他国内事例
 - 徳島大学病院におけるHDFSの利用
 - 名古屋大学医学部附属病院とベンダーによる共同研究プロジェクト

Copyright (C) 2013 Kozo Mizutani, Teikyo University

2. 効率的なデータ活用に向けて

◆ 医療情報分野でHadoopをどう使う？

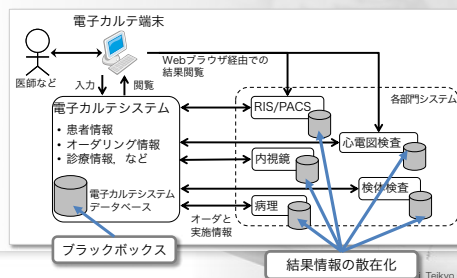
□ 特に病院情報システムにおいて

- いかに医療情報データを集積するか？
- 集積したデータをいかに分析するか？
- 分析結果をいかに医療へ役立てるか？

Copyright (C) 2013 Kazuo Mizutani, Teikyo University. 19

2. 効率的なデータ活用に向けて

□ 散在化する医療情報



Copyright (C) 2013 Kazuo Mizutani, Teikyo University. 20

2. 効率的なデータ活用に向けて

□ 帝京大学医学部附属病院でのアプローチ



帝京大学
医学部附属病院

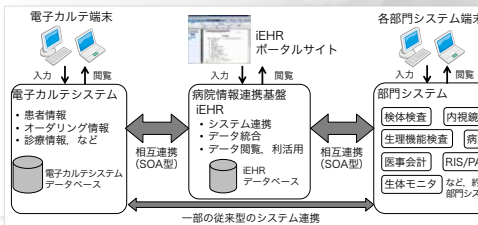
許可病床数：1154床
特定機能病院/救命救急センター/総合周産期母子医療センター/がん診療連携拠点病院/救急医療機関指定/災害拠点病院/臨床研修指定病院

Copyright (C) 2013 Kazuo Mizutani, Teikyo University. 21

2. 効率的なデータ活用に向けて

□ 帝京大学医学部附属病院でのアプローチ

- 2009年の新棟開院時からの新システム



Copyright (C) 2013 Kazuo Mizutani, Teikyo University. 22

2. 効率的なデータ活用に向けて

□ その他、課題

- 標準化への対応
 - DICOM, HL7
 - IHEのコネクタソンなど標準化相互接続
 - 病名、医学用語などの標準化など
- 紙文書の取扱い（電子化）
- UI統合型DBの利便性
 - 各科専門的な情報の蓄積など

Copyright (C) 2013 Kazuo Mizutani, Teikyo University. 23

2. 効率的なデータ活用に向けて

◆ 集積したデータをいかに分析するか？

□ 大小さまざまな医療情報 (大雑把な目安ですが)

- 医療画像
 - 断層画像、動画、3次元画像の場合はそれ以上
- 検体検査の結果：数KB～数MB/検査
- 定型フォームの値：数KB～数MB/頁
- オータ情報：数十バイト～数KB/件
- 生体モニタデータ：数十バイト～数KB/件など

情報粒度も様々 前提条件も様々

Copyright (C) 2013 Kazuo Mizutani, Teikyo University. 24

2. 効率的なデータ活用に向けて

- ◆ 集積したデータをいかに分析するか？
 - 分析のためのシステムアーキテクチャ
 - Hadoopの導入は有効的な分析基盤の1つ
 - 他、HBase、インメモリDBの活用など
 - 分析モデルの必要性
 - 自然言語処理
 - 臨床医学オントロジー
 - ▶ 東京大学を中心としたプロジェクト
 - 蓄積されたデータからのモデルの創出など。

Copyright (C) 2013 Koza, Mizutani, Teikyo University. 25

2. 効率的なデータ活用に向けて

- ◆ 今後何が必要か？
 - 1つの医療施設に限った範囲であっても技術的な課題は多い
 - 医療施設を超えて医療情報を蓄積・活用できる仕組みも必要

Copyright (C) 2013 Koza, Mizutani, Teikyo University. 26

2. 効率的なデータ活用に向けて

- ◆ 本格化しつつある地域医療連携
 - 標準化、セキュリティなど地域連携基盤の整備は相当程度進んできた
 - IDやプライバシーの確保などの課題解決が求められる
 - 政府主導のプロジェクト
 - 経済産業省 平成22年度「医療情報化促進事業」
 - 総務省 平成23年度「健康情報活用基盤構築事業」(日本版HER)

Copyright (C) 2013 Koza, Mizutani, Teikyo University. 27

3. 生体モニタデータへの適用

- ◆ Hadoopによる生体モニタデータの類似性分析
 - 手術中に使用する生体モニタからの計測値を定量的に扱う方法を検討するための基礎的研究。
 - 医療者は手術中のバイタルをどのように判断して、次のアクションに反映しているのか？
 - 仮説：教科書の知識と過去に経験した症例のバイタル変化を比較して判断している。

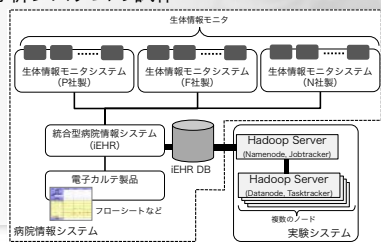


- どのように比較しているのか？
- バイタルが似ているとは？
- そもそも定量的に扱えるか？

Copyright (C) 2013 Koza, Mizutani, Teikyo University. 28

3. 生体モニタデータへの適用

- ◆ Hadoopによる類似性分析
 - 分析システムの試作



Copyright (C) 2013 Koza, Mizutani, Teikyo University. 29

3. 生体モニタデータへの適用

- ◆ Hadoopによる類似性分析
 - 2010年の手術事例
 - 5999症例
 - (手術スケジューリングシステムでの症例数)
 - 手術中の生体モニタデータのレコード数
 - 11,051,903レコード
 - 約398Mバイト
 - (IEHRからエクスポート時のテキストデータ量)

Copyright (C) 2013 Koza, Mizutani, Teikyo University. 30

3. 生体モニタデータへの適用

- ◆ Hadoopによる類似性分析
 - 2010年の手術事例～美データの例

| | | | | |
|---------------------|------|-------|-----|------|
| 2010-01-05T08:54:00 | OR13 | HR | 75 | bpm |
| 2010-01-05T08:55:00 | OR13 | HR | 72 | bpm |
| 2010-01-05T08:56:00 | OR13 | HR | 73 | bpm |
| 2010-01-05T08:57:00 | OR13 | HR | 70 | bpm |
| 2010-01-05T08:58:00 | OR13 | HR | 71 | bpm |
| 2010-01-05T08:59:00 | OR13 | HR | 70 | bpm |
| 2010-01-05T09:00:00 | OR13 | HR | 71 | bpm |
| 2010-01-05T09:01:00 | OR13 | HR | 67 | bpm |
| 2010-01-05T09:02:00 | OR13 | HR | 65 | bpm |
| 2010-01-05T09:00:00 | OR13 | NBPd | 70 | mmHg |
| 2010-01-05T09:00:00 | OR13 | NBPm | 88 | mmHg |
| 2010-01-05T09:00:00 | OR13 | NBPs | 129 | mmHg |
| 2010-01-05T08:54:00 | OR13 | Pulse | 76 | bpm |
| 2010-01-05T08:55:00 | OR13 | Pulse | 72 | bpm |
| 2010-01-05T08:56:00 | OR13 | Pulse | 72 | bpm |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

3. 生体モニタデータへの適用

- ◆ Hadoopによる類似性分析

- 症例ごとのデータ
 - 約5900症例分
- データの前処理
 - 約5900のタスク
 - 欠測への対応
 - 単位の考慮, など
 - 計測項目名をキーにOutput
- 分析処理 (複数のアルゴリズム)
 - サーバの数のタスク

3. 生体モニタデータへの適用

- ◆ Hadoopによる類似性分析
 - 結果データの例 (実際はバイナリ形式)

```

HR 20100107083000009879064 20100126140000009986898 101 301 501 10.244
HR 20100107083000009879064 999999999999999999999999 101 301 501 8.737
HR 20100107083000009879064 20100108160000009840052 101 301 501 9.817
HR 20100107083000009879064 20100107083000009819151 101 301 501 16.174
HR 20100107083000009879064 201001151645000007905505 101 301 501 20.777
HR 20100107083000009879064 201001290915000007246625 101 301 501 10.755
HR 20100107083000009879064 999999999999999999999999 101 301 501 8.718
HR 20100107083000009879064 20100121110000008320543 101 301 501 15.74
    
```

- 2010年 約5900症例分
- HRの場合約1.34 GB
- 12,422,620行

3. 生体モニタデータへの適用

- ◆ Hadoopによる類似性分析
 - 結果データの例

3. 生体モニタデータへの適用

- ◆ Hadoopによる類似性分析
 - 計測項目間での類似度の散布図例

- ✓ 類似度分析法: 1分間の変化量に対するユークリッド距離
- ✓ 対象データの期間: 手術開始後, 最大30分間

3. 生体モニタデータへの適用

- ◆ Hadoopによる類似性分析
 - 類似度のクラスター分析例

3. 生体モニタデータへの適用

◆ Hadoopによる類似性分析

- ✓ 水谷晃三, 澤智博, “分散処理フレームワーク”Hadoop”を用いた生体情報モニタデータ活用のための一方式, “第31回医療情報学連合大会論文集, 2-E-3-2, pp.541-542, 2011. (若手奨励賞受賞)
- ✓ 水谷晃三, 澤智博, “Hadoopを活用した生体モニタデータの類似性分析システム, “第16回医療情報学春期学術大会抄録集, PB-2-2-5, pp.184-185, 2012.
- ✓ 水谷晃三, 澤智博: テンプレートデータを用いた生体モニタデータの類似性分析に関する研究, 第32回医療情報学連合大会, 2-C-3-4.
- ✓ (社) 医療情報システム創成機構: 医療情報データ処理におけるHadoop導入の試み
<http://www.ahisi.org/reports/hadoop/index.html>

Copyright (C) 2013 Kozo Mizutani, Teikyo University. 37

4. おわりに

◆ Hadoopの期待大!

- 医療情報分野でHadoopが活躍できる場は多い!
- 医療施設を超えたデータ活用基盤が実現されれば, 益々重要な存在になるかも!
- 医療情報分野に特化したHadoop関連ソリューションの登場に期待!

Copyright (C) 2013 Kozo Mizutani, Teikyo University. 38

ご静聴ありがとうございました

TEIKYO

Copyright (C) 2013 Kozo Mizutani, Teikyo University. 39