

## 遠隔医療データのエラー検出に関する研究 Remote Healthcare Data Collection and Anomaly Detection

梶栗 芳夫†      甲斐 瑛子†      アシル アハメッド† ‡  
Yoshio Kajikuri      Eiko Kai      Ashir Ahmed  
ラフクル イスラム マルフ ‡      金子 邦彦†      福田 晃†  
Rafiqul Islam Maruf      Kunihiko Kaneko      Akira Fukuda

### 1. はじめに

医療データには性別・年齢・体型などにより個人差があるため、エラーを検知することが困難である。

現在、九大とグラミンは遠隔医療システムの研究開発を目指しており、農村部で集団健診を行い、都市部にいる医者が診断する形式で実証実験を行っている。

現場では、医療補助スタッフが村人の健診を行い、健診機器からデータを機械や手で電子化し、遠隔にいる医者に共有している。そのため健診の際に人的な入力ミスや計測誤差によりエラーが起きている。エラーを含むデータからは、正確な診断ができず、データの確認作業に大きな手間がかかっている。また、医者がエラーに気づき医療補助スタッフに再健診を求めることがあるため、負担となっている。

本研究では、遠隔医療データのエラー検出に向けて、遠隔医療データの特徴と既存の異常検出方法を調べ、ポータブルヘルスクリニックのデータを用いて実験する事で、エラー検出を行った。今回は、遠隔医療実証実験のデータから身長・体重のエラー検出とその結果について考察を行った。マハラノビス平方距離を利用して100個の異常値を抽出した。その異常値を実験現場に問い合わせる、および手書きのデータと照らし合わせることでエラーであるか否かを確認し、9個のエラーを検出した。

### 2. 遠隔医療データ

本章では研究背景と目的として、遠隔医療データの概要、ポータブルヘルスクリニックシステムの遠隔医療データについて述べる。

#### 2.1 遠隔医療とは

遠隔医療とは、通信技術を活用した健康増進、医療、介護に資する行為のことを指す。遠隔医療の形式には、主として主治医と専門医の間で実施される医療従事者間で行われるものと、医者から患者に対して行われる2つの形式があるが、本研究では医者と患者の間の医療の形式に注目する。対象となる患者は、開発途上国の農村地域における近くに病院のない患者や、震災においてインフラが完全に復旧していない地域の患者が挙げられる。

特に発展途上国では、医療インフラや医者・看護師などの医療資源が乏しく、またその多くが都市部に集中しているため農村部では医療を受けることが困難である。住民が病気になった場合にも、適切な治療を受けるために都市部

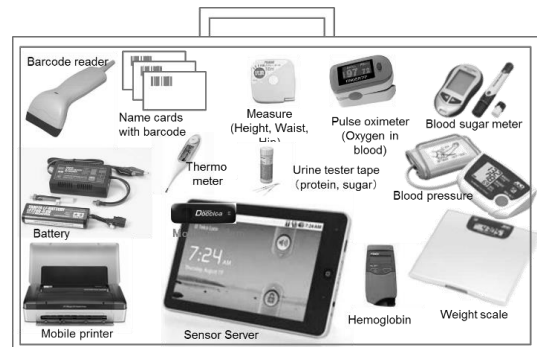


図1: ポータブルヘルスクリニック医療センサーセット

の病院に行くことはほとんどない。情報通信技術を利用し都市部の医者や医療サービスが行き届きにくい地域の患者を繋ぐことで、お金や時間の節約となる。

#### 2.2 ポータブルヘルスクリニックシステムにおける遠隔医療データ

ポータブルヘルスクリニックとは、現在、九大とグラミンは遠隔医療システムの研究開発を目指しており、農村部で集団健診を行い、都市部にいる医者が診断する形式で行っている実証実験のことである。2012年度は8527人に対して実証実験を行うと同時に、多種多様な遠隔医療データを取得している。ポータブルヘルスクリニックの遠隔医療データは以下の特徴がある。

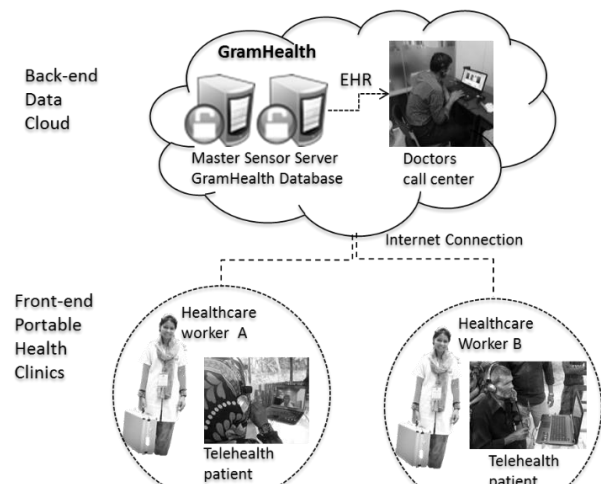


図2: ポータブルヘルスクリニックシステム全体図

†九州大学 Kyushu University, Japan

‡ Grameen Communications, Bangladesh

- ・医療データは医療補助スタッフによって取得される。現地では医療補助スタッフによって健診が行われデータ入力も医療補助スタッフによって行われる。
- ・一度に大勢の集団を健診してデータを入力する。  
より多くの人にサービスを利用できるように、集団で健診を行なっている。
- ・同じ機械を共有してデータを集める。  
先進国のように一人が医療健診機器を所有することが困難なので、同じ機械を共有して利用する。

### 2.3 データの種類と特徴

データの種類は主に、(1) 基本データ(2) 健診データ (3) 相談データ (4) 処方箋データ の4種類に分けることができる。それぞれ、構造化データ (数字) 半構造化データ (選択肢+自由記述) 非構造化データ (会話データ、テキストデータ) の特徴を持っている。データは GramHealth のデータベースに送られる。

#### ・問診データ

ポータブルヘルスクリニックでは利用者に対してまず、問診を実施する。問診では、既往歴の他、間食の有無や、1週間診にどのくらい運動しているか等の生活習慣を現地スタッフによる聞き取りにより調査する。また、名前や性別、住所、電話番号などの基本データも記入する。問診は選択式と自由記述式の2種類がある。

#### ・健診データ

問診の後、体重などの基本的な身体情報を始め、血圧、動脈酸素飽和度(SpO2)、血糖値などを測定する。健診ではそれぞれの測定項目に関して健診者を症状が軽い順に、健康、要注意、要治療、要緊急治療の4段階に分類する。この分類は国際的診断基準をもとに設定された基準(B-logic)に基づいている。最終的な判定結果は、各測定項目の内、もっとも症状が重い項目をその健診者の判断結果としている。

#### ・相談データ

首都ダッカのコールセンターには、医師が駐在しており、健診において「要治療」「要緊急治療」と判定された受診者に対して、医療相談を実施する。医師は、受診者の健診結果と Skype による遠隔相談を通じて、健康アドバイスを行う。このときの、患者の主張の要所をテキストで、医者との会話も音声データとして記録している。これらを相談データとする。

#### ・処方箋データ

利用者には、ポータブルヘルスクリニックの最後に処方箋をもらう。これには医者のアドバイスや必要となる薬名が含まれている。これを処方箋データという。健診データとともにモバイルプリンターで印刷され受信者に渡される。

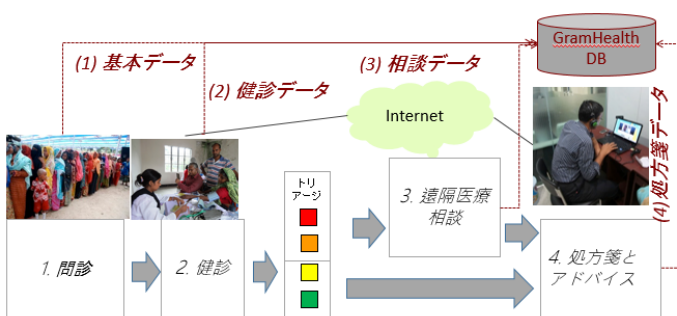


図3 遠隔医療データ収集方法

### 2.4 データの入力方法

それぞれのデータの入力方法は以下の通りである。

#### ・基本データ

医療補助スタッフは基本情報登録画面で、健診者の名前、性別などを登録する。それぞれの健診者には ID を指定できる。問診内容を医療補助スタッフが聞き取り記入する。それをノート PC に打ち込む。

#### ・健診データ

健診データはタブレットで専用のアプリケーションソフトを利用して入力される。医療機器とタブレットは Body Area Network (BAN) で接続されており、計測値が自動入力される。BAN に接続されていないものについては、スタッフが手動入力することが可能である。また同じ器具でも BAN 対応のものも非対応のものが存在する。またタブレットがない場合は、専用用紙に記入して後に医療補助スタッフによってノート PC に入力される。

#### ・相談データ

相談データにおける主訴情報は医者によって記録される。

#### ・処方箋データ

処方箋はコールセンターの医者によって入力される。薬はプルダウンメニューから選べるようになっており、容量は医者が追加で入力する。また患者へのアドバイスをテキストで自由に入力できる。

## 3. 遠隔医療データの課題

遠隔医療では医療データを計測する人とデータを診る人と医者が離れている。そのため医療データを正確に測ることが重要である。現在、ポータブルヘルスクリニックでは、一定のレンジをとり入力されたデータを確認している。

### 3.1 遠隔医療データのエラー

基本データと健診データについて、入力方法に関して、実験の実施箇所によって使用する医療機器が違うケースがあり、自動やマニュアルのどちらの場合も考える項目もある。間違いが起こる原因は大きく2つに分けることができる。

#### ・人に起因するもの

#### ・機器に起因するもの

人に起因するものとして、マニュアル入力による、読み取り間違い、記入間違い、記入ミスが考えられる。また医療補助スタッフの技術不足による計測方法の間違いも考えられる。故意の有無に関わらず間違いは起こる。機器に起因するものとして、機器の不具合が考えられる。

### 3.2 遠隔医療データのエラーが及ぼす影響

医療では医療データを計測する人とデータを診る人と医者が離れている。医者はその医療データを基に健康アドバイスをする。医療データが誤ると適切な判断ができない。またポータブルヘルスクリニックにおいては、測定結果を基に行われる b-logic によるトリアージの結果が誤ったものになり、実際は要治療の健診者を健康に分類してしまう恐れがある。このため、本当は治療が必要な健診者が治療を受けられない事態が起こりえる。現地では、データに疑問をもった医者が現場の医療補助スタ

ップに再健診を要請するケースもあり、貴重な医師の時間が奪われている。さらに現在、収集されたデータを分析・活用し、現地で実験を行う医療補助スタッフ支援システム構築をめざす研究も行われているためデータが誤っていれば分析に影響する。

	Green (Normal)	Yellow (caution)	Orange (affected)	Red (Emergency)
Waist	Male	≥90cm		
	Female	≥80cm		
Waist/Hip Ratio	Male	≥0.90		
	Female	≥0.85		
Body Mass Index(BMI)	<25	25≤ <30	30≤ <35	35≤
Blood Pressure (mmHg)	<130	130≤ <140	140≤ <180	180≤
	<85	85≤ <190	90≤ <110	110≤
Fasting Blood Sugar (FBS)	<100mg/dl	100≤ <126	≥126mg/dl	≥126mg/dl X 2times
Postprandial Blood Sugar (PPS)	<140mg/dl	140≤ <200	≥200mg/dl	≥200mg/dl X 2times
Urine Protein	- + =		≥+	
Urine Sugar	- + =	≥ I		
Urobilinogen	Normal or ±		Positive or +	
Pulse Ratio	80≤ <100	50≤ <60 100≤ <120	<50 120≤	
	Arrythmia	None	+	
Smoking	None	+		
Skin lesion	None		+	
Body Temperature	≤37°C	98.6 F ≤ <99.5 F <37.5°C	37°C ≤ 99.5 F ≤ (37.5°C ≤)	
SpO2	≥96%	93≤ <96	90≤ <93	<90%
Hemoglobin	≥12g/dl	10≤ <12g/dl	8≤ <10g/dl	<8g/dl

表 1 B-Logic によるトリアージ

#### 4. 実験

本研究では、マハラノビス平方距離を用いた異常検出を行った。そのうえで、異常値がエラーであるのかを検証した。

##### 4.1 異常値とエラー

本研究で扱う異常とは、遠隔医療データにおける項目に関して期待される傾向と一致しないものを指す。例えば、身長データの入力欄に 300cm の入力があった場合、これは明らかな異常値とみなす。次に、身長が 195cm との入力があった場合、この単体の数値だけでは異常値とはみなすことができないが、これと同時に体重のデータが 45kg というデータの入力があった場合、この数値が異常値である確立が高くなる。このようにして、明らかに異常値と定義できない場合においても、例えば身長や体重などに複数の条件を合わせて期待される傾向を求め、正常と異常の境界を定める必要がある。エラーとは正解の値と間違っていることを意味する。異常値は予測されるものであり、必ずしもエラーではない。特に医療データでは個人による体形の違いや、健康状態があるために、必ずしも統計的な異常値の境界線が適切であるとは言えない場合が多数あり、これを考慮しなければならない。同様に実際に起こっているエラーがすべて異常値として検出される訳ではなく、正常な集団の中に属すると定義されたデータの中にもエラーは存在する。

##### 4.2 実験データ

基本データと健診データをあわせたもの医療データを利用する。データはポータブルヘルスクリニックのデータから取り出した raw データと人の手によって確認済みの検査済みデータがある。raw データは、2013 年 11 月 13 日までの全データである。18279 回分のデータである。再診の利用者も含んでいる。検査済みデータは昨年実施された

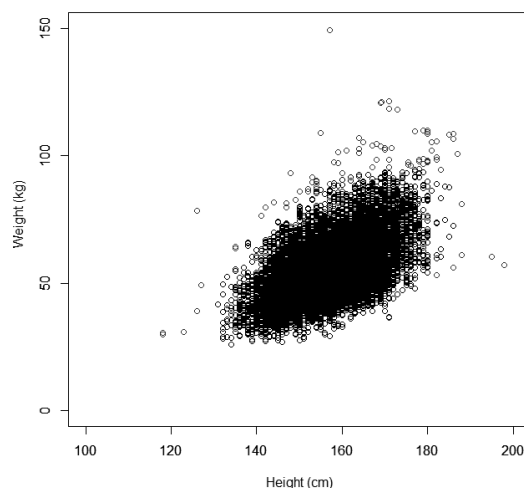


図 4 身長と体重の分布

実証実験における初回健診の 8527 人分のデータである。

データには、個人 ID、名前、性別、住所、携帯番号、健診日時、身長、体重、BMI、ウエスト、ヒップ、ウエストヒップ比、体温、血糖値、最高血圧、最低血圧、尿糖、尿タンパク、尿ウロビリノーゲン、血中酸素濃度、心拍数、不整脈かの判定、B-logic による診断の結果が含まれている。

##### 4.3 マハラノビス平方距離による異常値の検出

本研究では、マハラノビス平方距離を用いた異常検出を行った。実験は統計解析につかわれるオープンソース・フリーソフトウェアの R 言語を用いて行った。マハラノビス平方距離が大きい順に 100 個のデータを異常値としてエラーであるかの確認を行った。図 3 に身長と体重のプロットを示す。灰色のプロットが今回異常値として扱うものである。

#### 5. 実験結果

エラーの確認は、グラミンに問い合わせる方法と、手書きの健診情報と照らしあわせることで行った。異常値 100 個中、エラーであったのは 9 個であった。その結果を表 1 に示す。No1、No2 のように数字を誤って追加しまったエラーや No 4、5 の様に前後の数字を打ち間違えているエラーが確認できた。

No	身長(cm)	体重(kg)	間違ってる場所	正解
1	178	97.8	体重	79
3	157	149	体重	49
4	126	78.35	身長	162
5	195	60.4	身長	159
6	171	36.4	身長	162
7	187	100.7	身長	182
8	171	36.75	身長	167
9	186	56.2	身長	180

表 2 検出されたエラー

## 6. まとめ

本論文では、開発途上国における遠隔医療データのエラー検出に向けて、医療データの取得方法と構造を述べた。特に、遠隔医療データのマニュアル入力におけるエラーの可能性に着目し、マハラノビス平方距離を利用して異常検知を行った。その異常値からエラーの確認をグラミンに問い合わせることで行った。その結果、100 個の異常値の内、9 個がエラーであることが分かった。本論文では異常値の中からエラーを検知したが、正常値と判断したデータ中にもエラーが含まれていると考えられるため今後の課題とする。

### 参考文献

- [1] Ashir Ahmed, Sozo Inoue, Eiko Kai, Naoki Nakashima, Yasunobu Nohara, “ Portable Health Clinic: A pervasive way to serve the unreached community for preventive health- care ”, International Conference on Human Computer Intreraction, 2013
- [2] 野原康伸, Zahidu Ripon, Rafiqul Islam, Partha Ghosh, 井上創造, Ashir Ahmed, 中島直樹, “ 途上国における予防医療を実現するポータブルヘルスクリニックシステム の構築 ”, 第 39 回情報処理研究会 UBI 研究発表会, 2013
- [3] A. Anny Leema, Dr. Hemalatha. M, “ Anomaly Dtection and Elimination Algorithm for RFID Data ”, International Journal of Computer Applications, July 2012
- [4] 豊田秀樹編著, “データマイニング入門 -R で学ぶ最新データ解析”, 2008