

インフルエンザ感染モデルの提案と FCMによる各国・地域別エイズ実態評価

025208 高橋 知幸 025209 千原 晋平 025212 春山 秀幸

指導教官 佐藤 美佳

概要

1980年代には予防接種の普及や抗生物質療法の進歩により、感染症に対するリスクが軽減されてきた。しかし近年、インフルエンザの大流行や世界的な HIV 感染者及びエイズ患者の増加などによって感染症に対するリスクの重要性が再認識されるようになってきている。本論文ではインフルエンザ感染者数のシュミレーションを行う感染モデルを提案し、実データとの比較を行った。また統計データより FCM を用いてクラスタリングを行い、各国・地域別のエイズの実態について考察を行った。

1 はじめに

20世紀の医療のめざましい進歩や衛生環境の改善により感染症は減少し、それに対するリスクも軽減されてきた。しかし近年、エイズやエボラ出血熱に代表される新しい感染症が出現しており、著しく増加している。また結核などのように一旦は完全に制圧されていたと思われていた感染症の流行が、再び見られるようになってきている。特に最近の国際化、ボーダーレス化により地球的規模での対策が必要とされており感染症に対するリスクが再び重要視されるようになってきている。

本研究では代表的な感染症であるインフルエンザとエイズについて、それぞれ異なる立場から感染症に対するリスクについて分析を行った。インフルエンザを対象としては、患者が発生した場合の増加傾向を予測することを目的に、患者数の推移モデルを提案し、シュミレーションを行った。そしてその実験結果と実データとの比較を行い、提案したモデルについて考察を行った。エイズを対象としては、世界各国・地域別の患者数の分布を示す統計データを元に、クラスタリング手法の1つである FCM を用いることにより、各国・地域別のエイズの実態について考察を行った。

2 インフルエンザとは

2.1 インフルエンザの歴史

インフルエンザは時として多大な損害をもたらす。1890年(明治23年)にはアジアかぜが世界的に大流行。1918年には、スペインかぜが世界各地で猛威をふるい、全世界の罹患者数6億、死亡者は2,000-4,000万人とも推定されている。日本には大正8-9(1919-1920)年の冬に流行が持ち込まれ、罹患者は2,300万人、死者は38万人に及んだといわれる。

2.2 インフルエンザウイルス

インフルエンザウイルスはウイルス粒子内の核蛋白複合体の抗原性の違いから、A・B・Cの3型に分けられ、このうち流行的な広がりを見せるのはA型とB型である。A型ウイルス粒子表面には赤血球凝集素(HA)とノイラミニダーゼ(NA)という糖蛋白があり、HAには15の亜型が、NAには9つの亜型がある。これらは様々な組み合わせをして、ヒト以外にもブタやトリ、その他の宿主に広く分布している。

ウイルスの表面にあるHAとNAは、同一の亜型内でわずかな抗原性を毎年のように変化させるため、A型インフルエンザは巧みにヒトの免疫機構から逃れ、流行し続ける。これを連続抗原変異(antigenic drift)または小変異という。いわばマイナーモデルチェンジである。連続抗原変異によりウイルスの抗原性の変化が大きくなれば、A型インフルエンザ感染を以前に受け、免疫がある人でも、再び別のA型インフルエンザの感染を受けることになる。その抗原性に差があるほど、感染を受けやすく、また発症したときの症状も強くなる。そしてウイルスは生き延びる。

さらにA型は数年から数10年単位、突然別の亜型に取って代わることがある。これを不連続抗原変異(antigenic shift)または大変異という。インフルエンザウイルスのフルモデルチェンジで、新型インフルエンザウイルスの登場である。人々は新型に対する抗体はないため、大流行

となり、インフルエンザウイルスはさらに息をふきかえして生き延びる。

2.3 インフルエンザの症状と予防

日本のインフルエンザは、毎年11月下旬から12月上旬頃に発生が始まり、翌年の1-3月頃にその数が増加、4-5月にかけて減少していくというパターンであるが、流行の程度とピークの時期はその年によって異なる。

インフルエンザは発熱・頭痛・全身の倦怠感・筋関節痛などが突然現われ、咳・鼻汁などがこれに続き、約1週間で軽快するのが典型的な症状である。その他のいわゆるかぜ症候群に比べて全身症状が強いのが特徴である。

インフルエンザシーズンに、インフルエンザ流行に関連する肺炎死亡数は人口10万人あたり10人を越え(96/97, 98/99シーズン)、そのほとんどが65歳以上の高齢者であった。インフルエンザに関連すると考えられる脳炎・脳症で死亡した子どもたちは、年間100-200人に及ぶ。

インフルエンザに対して科学的な予防方法として世界的に認められているものは、現行のインフルエンザHAワクチンである。発病をほぼ確実に阻止するほどの効果はないが、高熱などの症状を軽くし、合併症による入院や死亡を減らすことができる。日本では、毎年インフルエンザシーズンの終わり頃にWHOからの情報および日本国内の流行情報などに基づいて、次シーズンのワクチン製造株が選定され、製造にとりかかる。現在はA型のH3N2とH1N1およびB型の3種のインフルエンザウイルスが世界中で共通した流行株となっているので、原則としてインフルエンザワクチンはこの3種類の混合ワクチンとなっている。

死亡率の減少などとともに、次第に「インフルエンザはかぜの一種でたいしたことはない」という認識が日本では広まってしまったが、決してそうではなく、国内でも地球的規模で見ても、インフルエンザは十分な警戒と理解が必要な疾患である。流行に伴う個人的・社会的損失はたいへん大きい。また新型インフルエンザウイルスの出現は必至である。これに対する警戒を怠ってはならない。[2][3]

3 エイズとは

エイズ(後天性免疫不全症候群)とはHIV(ヒト免疫不全ウイルス)が原因で、人間の免疫システムが損傷を受け、病原体に対する抵抗力が弱まることによって発症する様々な病気の総称である。HIVに感染すると平均10年の潜伏期間の後に、免疫力の低下に伴う感染症や悪性腫瘍、神経症状などの様々な症状が現れる。このよ

うにエイズが発症すると死亡率が高く、現在のところ完全に治療する方法は発見されていない。

3.1 HIVの感染経路

HIVの主要な感染経路は以下の3つである。

1. 性行為感染

HIV感染者と性行為を行った場合、精液・膣分泌液・血液中に含まれるHIVが粘膜をとって血液中に入り込み、HIVに感染する可能性がある。また、性器に傷がある場合は、HIVが傷口を通じて血管の中に入る可能性が大きくなるため、感染の可能性は大きくなる。現在、最も多いHIV感染原因がこの性行為による感染である。但し、コンドーム等を正しく使用することにより、感染を未然に防ぐことが可能である。

2. 母子感染

HIVに感染している母親が妊娠・出産時に子宮内や産道で胎児が感染する可能性がある。また出産後も母親の母乳から感染する可能性もある。但しこの場合においても、出産時に帝王切開を行うことや、母乳の代わりに粉ミルクを使用することによって、感染の可能性を低下させることが可能である。

3. 血液媒介感染

HIVに感染している麻薬使用者が使った注射器を使って麻薬を注射した場合、針や注射器中についていた感染者の血液が血管にはいることによってHIVに感染する可能性がある。また、HIV感染直後の6~8週間は抗体検査を行っても感染が検出できないため、感染者がその期間中に献血を行った場合、その血液が輸血にしようされることにより、感染の可能性はある。

3.2 HIV感染の予防

以上のいずれの感染経路についても、正確な知識と予防行動を心がけることにより、感染を未然に防ぐことが可能である。現在、エイズ問題を悪化させる最大の原因として、HIV・エイズに対する無知や偏見があげられる。エイズの恐ろしさを知らなかったり、偏見を恐れて検査やカウンセリングを受けないために、さらに感染が広がるという悪循環が起こっている。従ってエイズ問題においては、正しい知識の普及や予防行動の啓発が非常に重要になっている。

4 インフルエンザ感染者数のシミュレーション

本章では、本研究において提案したインフルエンザ感染者数の推移モデルについて述べ、実データを用いたシミュレーションによる実験結果から、その妥当性について検討を行う。シミュレーションを行うプログラムはC言語で実装した。

まず、インフルエンザ感染者数のように、流行が始まると指数関数的に増えて、流行が過ぎれば減少に転じるようなモデルを提案する。病気の指数関数的増加を取り扱うモデルはすでに提案されている [1] が、治癒者がいることは考慮されておらず、感染者数は増える一方であるためインフルエンザなどの感染者数をシミュレートするには向かない。そこで、既存のモデルを治癒者も考慮して改良する。改良したモデルを用いてインフルエンザ感染者数の推移シミュレーションを行い、実データを近似するための適当なパラメータの設定と、提案したモデルの妥当性を検討する。

提案したモデルには感染率、治癒率といったパラメータがある。これらのパラメータを適当に設定すると、実データをよく近似できる。実験1では一定の感染率、治癒率を用いてインフルエンザ感染者数の推移をシミュレートする。このとき、感染率、治癒率を少しずつ変化させながら、実データを最もよく近似するパラメータを求める。

実験中は何度も感染者数の更新を行うが、実験2では感染者数の更新のたびに感染率、治癒率が変動することを許容して、より実データに近い結果を得ることを試みる。そして、感染率、治癒率がどのような変化をしているか考察する。

4.1 提案したモデルについて

病気の指数関数的増加を取り扱うモデルの一つとして、 $I(t)$ を時間 t 、人口 N における病気の感染者数、 α を感染率としたとき、時間が dt だけ変化する間の新たな感染者数 $\frac{dI(t)}{dt}$ を、

$$\frac{dI(t)}{dt} = \alpha I(t) \left[1 - \frac{I(t)}{N} \right] \quad (1)$$

のように表すものがある [1]。このモデルは以下のように導出される。まず、時間が dt だけ変化する間に、 $I(t)$ 人の感染者が $\alpha I(t)$ 人に病気を感染させる。 $\frac{I(t)}{N}$ はすでに感染している人の割合を表すので、新たな感染者数は

$$\alpha I(t) - \alpha I(t) \left[\frac{I(t)}{N} \right] \quad (2)$$

となる。これより、式 (1) は容易に導出できる。

このモデルを用いたときの $I(t)$ と $I(t + dt)$ との関係、時間 t と $I(t)$ との関係はそれぞれ図 1, 2 のようになる。図 1 は式 (1) における $I(t + dt)$ と $I(t)$ との関係をもとに $N = 10000$ としてグラフ表示したものである。 $I(t) = N$ のとき、 $I(t + dt) = I(t) = N$ となるので、感染者数 $I(t)$ が人口 N を超えることはない。また、図 2 からわかるように、 $I(t)$ は一様増加の関数になっている。そのため、エイズのように一度感染すると完全に治癒しない、つまり感染者数が減ることのない病気に対しては式 (1) によって感染者数のシミュレートが可能と思われる。しかし、インフルエンザのように流行が始まると感染者は増えるが、治癒者がいるために流行が過ぎれば感染者数が減少に転じようとする病気に対しては、式 (1) のモデルでは感染者数のシミュレートができない。そこで、感染者数の算出と同時に治癒者の算出も行い、さらに、同じシーズンでは一度感染した人は二度と感染しないという制約をつけることで感染者数のシミュレーションにおいて、感染者数のピークをむかえた後に感染者数が減少に転じるようなモデルを提案する。

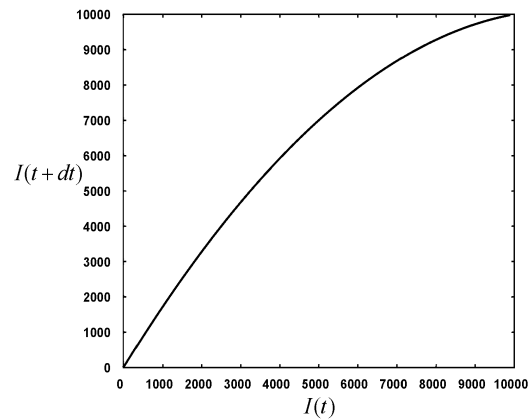


図 1: 式 (1) における $I(t + dt)$ と $I(t)$ との関係。

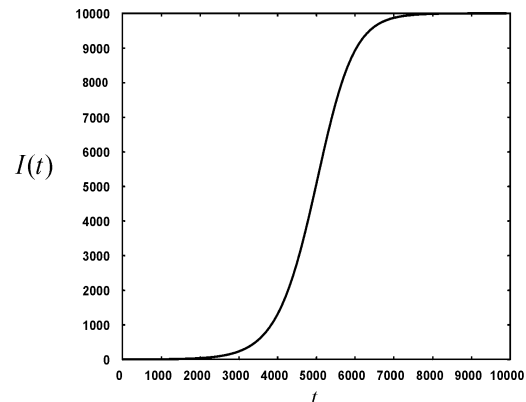


図 2: 式 (1) における時間 t と感染者数 $I(t)$ との関係。

表 1: 三重県のインフルエンザ患者報告数.

観測週	45	46	47	48	49	50	51
報告数	1	1	1	1	1	6	9
観測週	52	1	2	3	4	5	6
報告数	15	19	114	271	589	1064	1873
観測週	7	8	9	10	11	12	13
報告数	1533	2083	1827	1379	1011	608	292
観測週	14	15	16	17	18	19	20
報告数	126	64	51	28	9	4	0

今回用いた新たな感染者数は、 $c(t)$ を時間 t においてすでに治癒した人の数としたとき、

$$\frac{dI(t)}{dt} = \alpha I(t) \left[1 - \frac{I(t)}{N - c(t)} \right] \quad (3)$$

で与える。また、新たな治癒者数 $\frac{dc(t)}{dt}$ を

$$\frac{dc(t)}{dt} = \beta I(t) \quad (4)$$

とする。 β は治癒率で、 $\beta = 0$ のときは式 (1) と式 (3),(4) は同じになる。

α と β の値や、計算の時間間隔 dt を適切に選べば、実データの値を良く近似することができる。そこで、実際に患者数を集計したデータを基に、感染率、治癒率を変化させながら式 (3),(4) を用いてシミュレーションを行い、最も良く実データを近似することができる感染率 α 、治癒率 β を算出する。近似のよさは、実データと実験データとの間の平方 2 乗誤差の和を計算して評価する。

4.2 インフルエンザの実データ

インフルエンザの実データとしては、表 1 に示した 2001/2002 年の三重県における全定点からの患者報告数を用いる [4]。定点とは、インフルエンザ患者を確認したときに、その数を保健所に報告する義務を負う報告医療機関のことである。三重県内には 73 の定点が存在しており、毎週各定点医療機関から患者数が報告される。それを集計したものが表 1 である。

また、観測週が 45 からスタートしているが、これは 2001 年の第 45 週目という意味で、この表においては 11 月 5 日から 11 月 11 日までを指す。観測週が 1 というのは 2002 年の第 1 週目で 12 月 31 日から 1 月 6 日までとなっている。三重県では 2001 年の第 45 週にその年初めてのインフルエンザ患者が報告されている。そして、2002 年の第 20 週に報告数は 0 となった。つまり、初めて報告が

あった週から報告がなくなった週までの 28 週間分がまとめられている。

実験では表 1 のデータを用いてシミュレーションを行う。表 1 の患者報告数をグラフ表示すると、図 3 のようになる。患者数がある程度まで増えると患者数は減少に転じていることがわかる。

実験 1 では、表 1 の患者報告数を三重県のインフルエンザ感染者数と仮定して、式 (3),(4) を用いたシミュレーションを行う。ただし、県内には報告医療機関以外にも多くの医療機関が存在し、インフルエンザにかかっても症状が軽い人は病院に行かない場合もあるので、実際にはこれよりも多くの感染者がいると思われる。

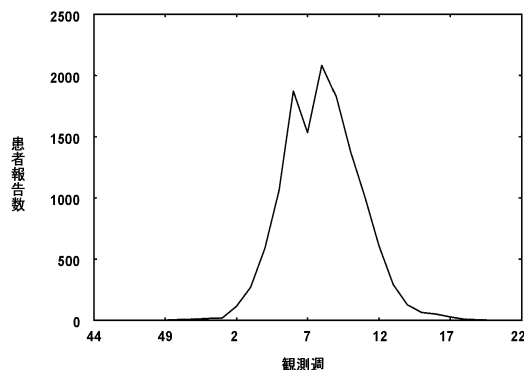


図 3: 三重県における 2001/2002 年の患者報告数のグラフ。横軸は観測週、縦軸は報告患者数を表す。

4.3 実験 1

式 (3),(4) の各パラメータは以下のように設定した。

- $N = 18600$: 三重県の人口は約 186 万人だが、患者報告数とスケールをあわせるために $\frac{1}{100}$ 倍した。
- $I(0) = 1$: 初期感染者数は 1 人。
- 感染者数の更新は一日 24 回、28 週間分行う。
- 感染率 α の初期値は 0.0、そして、これを 0.0005 ずつ増やしていく。
- 治癒率 β の初期値は 0.0、そして、これを 0.0005 ずつ増やしていく。

このようなパラメータの下で、感染率、治癒率を少しずつ変化させながら、最も良く表 1 の報告数を近似する感染率、治癒率を導出する。近似の良さは報告数のデータと、その時点で相当する実験データとの平方 2 乗誤差の和をもとに評価する。

実験1の結果を図4,5に示す。図4は、感染率、治癒率の各値における誤差を等高線表示したものである。誤差が小さくなるにつれて色が黒に近づいていく。感染率0.0125, 治癒率0.0090のところでは誤差が最も小さくなった(+印)。つまり、感染率、治癒率が実験区間である28週間は常に一定であるとき、最も良い近似を与える感染率は0.0125, 治癒率は0.0090であるということがわかる。

また、図5は、式(1)のモデルと式(3),(4)のモデルを使って感染者数をシミュレートしたグラフを比較したものである。今回提案した式(3),(4)のモデルでは適切にパラメータを調整することによって実データを近似できるが、式(1)のモデルでは実データをうまく近似できないと言える。図5におけるの実線は感染率を0.0125, 治癒率を0.0090として式(3),(4)を用いて感染者数の推移を求めたもの。実データをよく近似していることがわかる。

の破線は感染率を0.0125のまま治癒率を0.0にしたもの、つまり式(1)において感染率をと同じく0.0125としたものであるが、感染者数は実データをはるかに上回ってしまっている。また、の破線は式(1)を使ったときに、最も良く実データを近似しているグラフである。感染率は0.0012となったが、実データと比べて感染者数が極端に少ない。式(1)を使うとこのような結果になってしまったのは、式(1)の単調増加の性質によるものである。

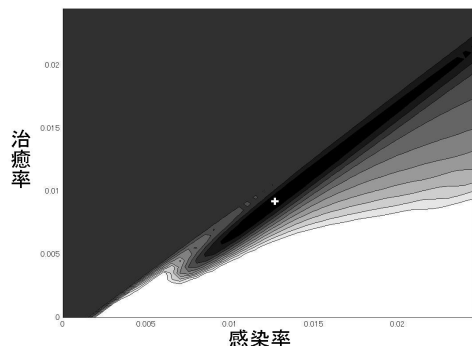


図4: 感染率、治癒率の各値における誤差を等高線表示したグラフ。

4.4 実験2

実験1においては感染率、治癒率は実験区間全体で一定であった。実験2では、実験1で得られた $\alpha = 0.0125$, $\beta = 0.0090$ を三重県における感染率、治癒率の真値とみなして、真値を中心として感染者数の更新のたびに感染率、治癒率も変化させることでより実データに近い結果を得ることを試みる。また、感染率、治癒率の変遷をグ

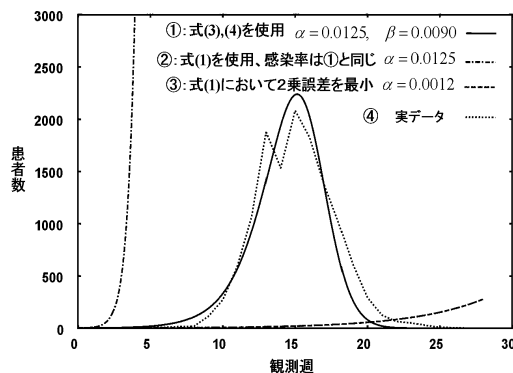


図5: 式(1)のモデルと式(3),(4)のモデルとの比較。

ラフにし、両者の関係を考察する。

実験2では、式(3),(4)の各パラメータは以下のように設定した。

- $N = 18600$: 実験1と同じく、三重県の人口を $\frac{1}{100}$ 倍している。
- $I(0) = 1$: 初期感染者数は1人。
- 感染者数の更新は一日24回, 28週間分行う。
- 感染率、治癒率は毎回変化させるので、 $\alpha(t)$, $\beta(t)$ とする。

感染率、治癒率は以下のように設定した。まず、式(3),(4)を変形すると、

$$I(t + \Delta t) = I(t) + \alpha(t)I(t) \left[1 - \frac{I(t)}{N - c(t)} \right] - \beta(t)I(t) \quad (5)$$

となる。ここで、実データにおける時点 $t + \Delta t$ での値を感染者数の目標値 I' としたとき、 $I(t + \Delta t)$ の値を感染者数の目標値 I' と等しくしたい。式(6)の $I(t + \Delta t)$ に I' を代入して変形すると、

$$I' - I(t) = \alpha(t)I(t) \left[1 - \frac{I(t)}{N - c(t)} \right] - \beta(t)I(t) \quad (6)$$

となる。 $I(t)$, N , $c(t)$, I' の値はすでにわかっており定数と考えることができるから、式(6)は $\alpha(t)$, $\beta(t)$ を変数とする式となる。ここで、感染率 $\alpha(t)$, 治癒率 $\beta(t)$ に実験1で得られた結果をもとに感染率の目標値 $A = 0.0125$, 治癒率の目標値 $B = 0.0090$ を設定し、

$$J(t) = (\alpha(t) - A)^2 + (\beta(t) - B)^2 \quad (7)$$

が最小になるように $\alpha(t)$, $\beta(t)$ を決めた。また、 $\alpha(t)$, $\beta(t)$ は共に正であるという制約も設けた。

さて、ここで問題となるのが感染者数の目標値 I' の取り方である。実データとしては各週の報告患者数がある

だけなので、実験区間の任意の時点から患者数の推定値を得るには工夫が必要になる。そこで、各週の患者数のデータを何らかの形で補間して、任意の時点において患者数の推定をすることができるようにする。本研究では各週の患者数データを3次スプライン関数で補間した。3次スプライン補間法とは、離散データを3次の区分的多項式で滑らかに補間するものである。これにより、実験区間の任意の時点で患者数の推定値を得ることができ、感染者数の目標値 I' の設定が可能になる。

近似のよさも、実験区間の任意の時点における患者数の推定値を得ることができるようになったので、感染者数の更新をするたびに感染者数の目標値とシミュレーション結果との平方2乗誤差の和を算出し、実験区間におけるそれらの積算を求めることで評価を行った。

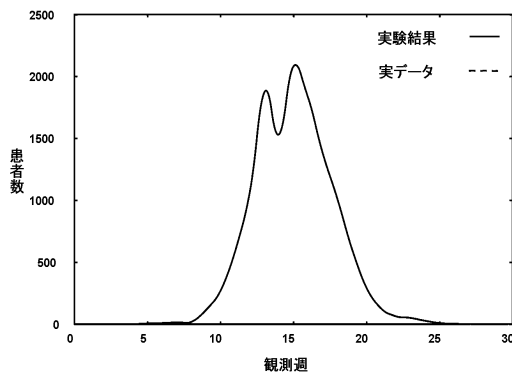


図 6: 感染率, 治癒率を毎回変えることを許したときの感染者数の推移.

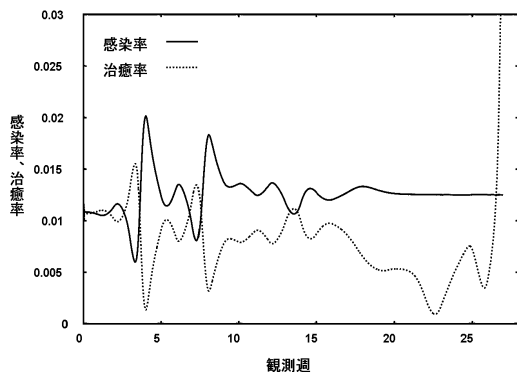


図 7: 感染率, 治癒率の推移.

結果を図 6,7 に示す。図 6 は実験 2 によって得られた感染者数の推移である。平方 2 乗誤差の和は無視できるほど小さくできた。実データを 3 次スプライン補間法で補間したもの、つまり感染者数の目標値とほとんど変わらない。図 7 は実験 2 における感染率 $\alpha(t)$, 治癒率 $\beta(t)$ の推移で、実線は感染率, 点線は治癒率を表す。流行初期

で感染率, 治癒率の波が大きくなっているのが確認できた。15 週前後の感染者数が増えてから一旦落ち込み, また増えるところでは, 治癒率が感染率を逆転していた。

4.5 本章のまとめと今後の課題

本章では, インフルエンザ感染者数の推移モデルを提案し, 実データを用いたシミュレーションにより, その妥当性を検討した。

実験 1 では, 感染率, 治癒率はそれぞれ一定であるという仮定の下で, 実データを最もよく近似する感染率, 治癒率を求めた。式 (1) のモデルではインフルエンザのような感染者数が一旦増えてもある時点を境に減少に転じるような病気に対しては感染者数のシミュレーションはできなかったが, それを改良した式 (3),(4) を提案し, 用いることで実データを近似することができた。

実験 2 では, 感染者数の更新ごとに異なる感染率, 治癒率を用いてもよいとする仮定の下でより実データに近いシミュレーション結果を得ることができた。

今後は, もし地域別, あるいは年度別のインフルエンザ感染者数のデータが得られれば, 地域別, 年度別, ウイルスの種類別のインフルエンザ感染の規模や特徴を考察できると思うので, データを探して実験したい。

5 クラスタリングを用いた各国・地域の HIV/AIDS 感染者の実態評価

今回, 各国・地域の HIV/AIDS 感染者の実態を評価するため, クラスタリングを行うことによって, 様々な側面から実態を考察することにした。なお, データは [5] より抜粋したデータを用いた。

なお, この資料は HIV/AIDS 感染者の推定データであるが, 各国によっては子供の患者数や死亡者数, 孤児の患者数などのデータが無い国があったため, まず今回データを扱うにあたって, 総患者数, 大人の患者, 女性の患者, 子供の患者, 死亡者数の 5 変数を使用した。なお, これらのデータも欠けている国々があったため, 全体的にこの 5 変数が提供されている国 49 ヶ国をリストアップし, HIV/AIDS の実態評価を行った。

5.1 FCM

ファジィクラスタリングは, 個体がクラスターに帰属する度合いにあいまいさを導入するという考えに基づいている。あいまいさの表現はファジィ理論による要素の集合への帰属度 (メンバーシップ) によって表される。

まず、個体 x_k がクラスター G_i に属する変数（帰属度）を u_{ik} とする。そして、変数 u_{ik} に 0 から 1 までの間の値を連続的に取る。つまり、各個体についてすべてのクラスターに対する帰属度を加えると 1 になるという正規化の条件を加える。式で表すと

$$M_f = \{(u_{ik}) : u_{ik} \in [0, 1], \sum_{i=1}^c u_{ik} = 1 \text{ for all } k\} \quad (8)$$

というようになる。

個体は p 次元ユークリッド空間の点 $x_k = (x_k^1, x_k^2, \dots, x_k^p)^T, k = 1, \dots, n$ で表す。また、クラスター中心 $v_i = (v_i^1, v_i^2, \dots, v_i^p)^T$ であり、これを $V = (v_1, \dots, v_c)$ とする。

そして FCM では次のような目的関数を考える。

$$J_{fcm}(U, V) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m d_{ik} \quad (9)$$

ここで、 m は $m > 1$ を満たすようなべき乗パラメータである。そして、 $d_{ik} = \|x_k - v_i\|^2$ とする。具体的なアルゴリズムは

STEP1

V' の初期値を与える。

STEP2

V' を固定して

$$\min_{U \in M_f} J(U, V')$$

を解き、最適解を U' とする。

STEP3

U' を固定して

$$\min_V J(U', V)$$

を解き、最適解を V' とする。

STEP4

解 (U', V') が収束すれば終了。そうでなければ $\bar{U} = U', \bar{V} = V'$ として STEP2 に戻る。そして、収束条件は

1. 帰属度 u_{ik} がある一定の大きさに収束したら終了。

$$\max_{i,k} |\bar{u}_{ik} - u'_{ik}| < \epsilon$$

2. クラスター重心 v_i がある一定の大きさに収束したら終了。

$$\max_{i \leq c} \|\bar{v}_i - v'_i\| < \epsilon$$

3. 目的関数 $J(eq.(9))$ が収束したら終了。

このどれかを用いて収束条件を決定する。

5.2 データの基準化

今回、各国・地域別に HIV/AIDS の推定感染者数をクラスタリングを行うことによりどのような特徴があるのかを分析した。しかし、データの項目には推定総患者数、大人、女性、子供、死亡者があるが、これらのデータを素直に使用しようとする各項目によって桁数がまったく異なり、当然、推定総患者数が桁数が多く、この推定総患者数にクラスタリング結果が左右されてしまうといったことが起きる。さらに、南アフリカなど HIV/AIDS の推定感染者数が非常に大きい国と香港など推定感染者数が非常に小さい国とのデータをそのまま使用すると、あるパラメータが突出した国にクラスター重心が引き寄せられ、小さい値のところのデータの構造をクラスタリングによって抽出することが出来ない状況が起こる。そこでまず各国々の推定総患者数に対する、大人、女性、子供、死亡者の割合 (%) を求め、さらに総患者数も含めてこれらを基準化し、クラスタリングを行った。なお、この基準化には以下の基準化の手法を用いた。

- x_{ia} : 元のパラメータ
- z_{ia} : 基準化されたパラメータ
- $i = 1, \dots, 48$: 国（個体）数
- $a = 1, \dots, n$: クラスター数

$$\bar{x}_a = \frac{\sum_{i=1}^n x_{ia}}{n} \quad (10)$$

ここで z_{ia} と σ_a は

$$z_{ia} = \frac{x_{ia} - \bar{x}_a}{\sigma_a} \quad (11)$$

$$\sigma_a = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ia} - \bar{x}_a)^2} \quad (12)$$

とした。

5.3 クラスタ数の決定

今回の AIDS/HIV の実態評価のクラスタリングを行うにあたって、どのクラスタ数を選択するかを、目的関数 J を使って決定した。本来クラスタ数は、より少ない数でデータの構造が解析出来るのが理想である。つまり、単にクラスタ数を増やしてクラスタリングを行っても、データの構造が大変複雑になり、手法の安全性を考慮するならばデータの構造をよく抽出した出来るだけ小さいクラスタ数を選択することが望ましい。この目的関数のグラフは、図 8, 9, 10 に示しておく。なお、4 変数（大

人, 女性, 子供, 死亡者) の場合においては, クラスタ数の選択が難しいため, クラスタ数 2, 3 それぞれの場合においてクラスタリングを行った. 他のものである. 減少量が少なくなったと考えられる, クラスタ数 3 で行った.

5.4 実験の概要

本実験では

1. 女性 (%) - 子供 (%) : クラスタ数 3
2. 総疾患患者数 (人) - 女性 (%) - 子供 (%) : クラスタ数 3
3. 大人 (%) - 女性 (%) - 子供 (%) - 死亡者 (%) : クラスタ数 2
4. 大人 (%) - 女性 (%) - 子供 (%) - 死亡者 (%) : クラスタ数 3

の 4 項目について国別にクラスタリングを行った. そして, クラスタ数の決定であるがこれは目的関数 J (eq.(9)) の減少が落ち着いたところでクラスタ数を決定している. クラスタ数を決定するに当たって, 利用したこの目的関数 J のクラスタ数による増減を表したグラフを後ろに添付しておく. なお, 今回使用した目的関数 J の式の中の m に関して, $m = 2$ を使用した.

なお, FCM についてはプログラムを作成し, クラスタリングを行った. また, 各地域別 (サハラ以南地域, 東アジア・太平洋地域, オーストラリア・ニュージーランド地域, 南・東南アジア地域, 東ヨーロッパ・中央アジア地域, 西ヨーロッパ地域, 北アフリカ・中東地域, 北アメリカ地域, カリブ海地域, 南アメリカ地域) についても同様にクラスタリングを行った.

5.5 結果と考察

5.5.1 国別データに対する結果

1. 女性 (%) - 子供 (%) : クラスタ数 3
クラスタリング結果を表 2 に示す. 図 1 1 はクラスタリングを行った結果得られた帰属度をグラフ化したものである. まず, クラスタからの観点で見ると, Cluster1 については, 女性の推定疾患患者割合が多く, 子供の推定疾患患者割合も多い集合である. 次に Cluster2 であるが, ここは推定女性疾患患者が少なく, さらに子供の推定疾患患者も少ない集合である. Cluster3 であるが, ここは若干女性の推定疾患患者割合が大きい, 子供の推定疾患患者割合が少ない集合である.
2. 総疾患患者数 (人) - 女性 (%) - 子供 (%) : クラスタ数 3

次に総疾患患者数 (人) - 女性 (%) - 子供 (%), クラスタ数 3 について考える. 結果は表 3 に示してある. 図 1 2 は各クラスタごとに帰属度をグラフ化したものである. これは先ほどの女性 (%) - 子供 (%) クラスタ数 3 に各国の総疾患患者数のパラメータを加えたものである.

Cluster1 については 2 ヶ国しか所属していないが, これは総疾患患者数のパラメータが大きく影響を及ぼしている. クラスタ中心を見ても分かるように, Cluster1 では総疾患患者数が際立って大きくなっている.

次に Cluster2 であるが, これは先ほどの Cluster1 より推定総疾患患者数が低くなり, 後は Cluster1 と同じような女性, 子供の推定疾患割合である. ここでは, Cluster1 ほど, 総疾患患者数が多くないのに, 女性の推定疾患患者割合が大きく, さらに子供の推定疾患患者割合も高い結果になっている.

次に Cluster3 であるが, これは全体的に推定総疾患患者数も女性, 子供の推定疾患割合も少ない国々の集合であると見られる.

3. 大人 (%) - 女性 (%) - 子供 (%) - 死亡者 (%) : クラスタ数 2
次に大人 (%) - 女性 (%) - 子供 (%) - 死亡者 (%) : クラスタ数 2 について考える. クラスタ中心から見ると, Cluster1 に関しては大人の推定疾患割合が大きい, 女性, 子供, 死亡者の割合は小さくなっている. Cluster2 に関しては大人の推定疾患割合が小さいにもかかわらず, 推定女性, 子供, 死亡者の割合が高めであるということが分かる. ここでの大きなポイントは, 大人の推定疾患患者割合である. つまり, Cluster2 では大人の疾患患者割合が小さく女性, 子供の疾患割合が高い.
4. 大人 (%) - 女性 (%) - 子供 (%) - 死亡者 (%) : クラスタ数 3

次に, 大人 (%) - 女性 (%) - 子供 (%) - 死亡者 (%) をクラスタ数 3 で検証してみた. まず, クラスタ中心から見ると, クラスタ数を増やしたことによって, 先ほどの Cluster1 が 2 つのクラスタ (Cluster1, Cluster3) に分かれている. Cluster2 に関しては, ほとんどといっていいほど中心座標が変化していない. これによって, 先ほどの Cluster1 が大人, 女性の推定疾患患者割合が高い Cluster1 と大人の推定疾患患者割合が高い国 Cluster 3 に分割されている.

5.5.2 地域別データに対する結果

地域別にクラスタリングを行った結果について述べる. 地域別データに対しても同様に 4 パターンに対してクラ

スタリングを行った。

1. 女性, 子供, クラスタ数 3

女性 (%) - 子供 (%) について, 地域別のデータに対してクラスタリングを行った。地域別データの場合では, 女性と子供の双方の推定疾患割合が高いクラスタと, 双方の割合が比較的低いクラスタ, 及び特に双方の割合が低いクラスタに分類された。

2. 総疾患患者数 (人) - 女性 (%) - 子供 (%) : クラスタ数 3

総患者数, 女性, 子供の 3 つの変数について地域別データに対してクラスタリングを行った。このデータに対しては総疾患患者数の単位が異なるため基準化を行っている。この場合も同様に女性と子供の推定疾患割合の程度によって, クラスタリングされている。Cluster2 には女性, 子供双方の推定疾患割合が高い地域が, Cluster3 には双方の割合が低く, 患者数も低い地域が属しており, Cluster1 にはその中間的な地域が属している。Cluster1 と Cluster2 には特に総患者数の違いは見られないが, 女性と子供の推定疾患割合の違いによって分類されているものと考えられる。

3. 大人 (%) - 女性 (%) - 子供 (%) - 死亡者 (%) : クラスタ数 2

大人 (%), 女性 (%), 子供 (%), 死亡者 (%) を変数としてクラスタ数 2 でクラスタリングを行った。この場合では, 大人の推定疾患割合はあまり関係なく, 女性, 子供, 死亡者の割合の程度によってクラスタリングされている。

4. 大人 (%) - 女性 (%) - 子供 (%) - 死亡者 (%) : クラスタ数 3

パターン 3 についてクラスタ数を 3 にしてクラスタリングを行った。この場合も他の 3 パターンと同様に, ほとんど同じ地域の組み合わせで分類されている。

6 まとめ

本研究では, 代表的な感染症であるインフルエンザとエイズを対象とし, それぞれ異なる立場から感染症に対するリスクの分析を行った。インフルエンザにおいては, 感染者数の推移モデルを提案し, シミュレーションを行い, 実データに近い結果を得ることができた。また, エイズでは, エイズ推定疾患患者の各国別・地域別の分布を示す統計データを元にクラスタリングを行った。実験結果より, 各国・地域別のエイズの実態の傾向を把握することができた。

参考文献

- [1] D.M. カーメン, D.M. ハッセンザール著, 中田俊彦訳, 「リスク解析学入門」, シュプリンガー・フェアラーク東京, 2001
- [2] 根路銘国昭, 「インフルエンザ大流行の謎」, 日本放送出版協会, 2001
- [3] 国立感染症予防センター, インフルエンザ総説
<http://idsc.nih.go.jp/others/inf-soron.html>
- [4] 三重県保健所管内別インフルエンザ患者報告数及び定点当たり患者数 (2001/02 シーズン)
<http://www.kenkou.pref.mie.jp/topic/inful/infhyou0102.htm>
- [5] Table of county-specific HIV/AIDS estimates and data, end 2001
<http://www.unaids.org/barcelona/presskit/barcelona%20report/table.html>
- [6] JCA-NET
<http://www.jca.apc.org/>
- [7] 宮本 定明: クラスタ分析入門, ファジイクラスタリングの理論と応用, 森北出版, 1999.

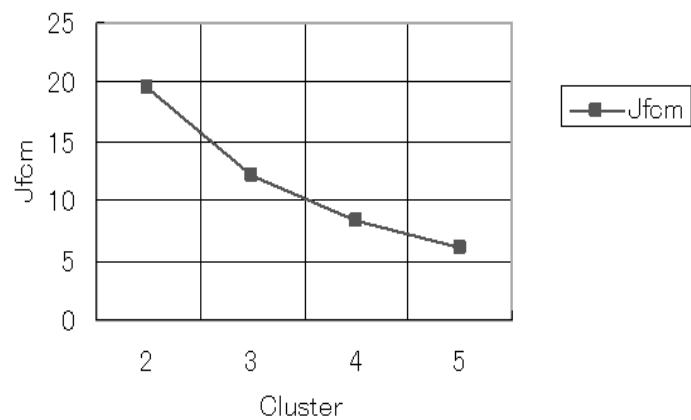


図8 目的関数の増減、変数(大人、子供、女性、死亡者)

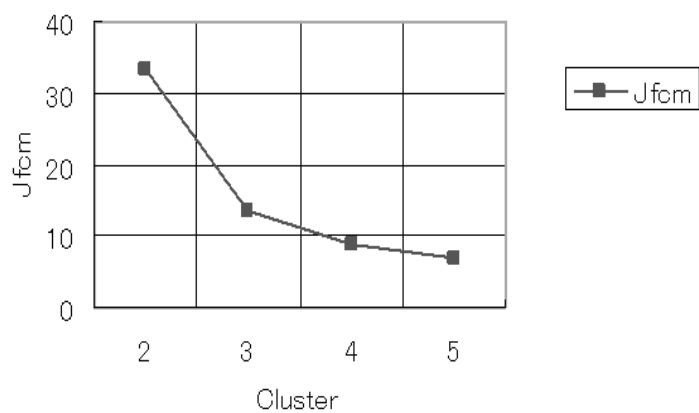


図9 目的関数の増減、変数(患者数、女性、子供)

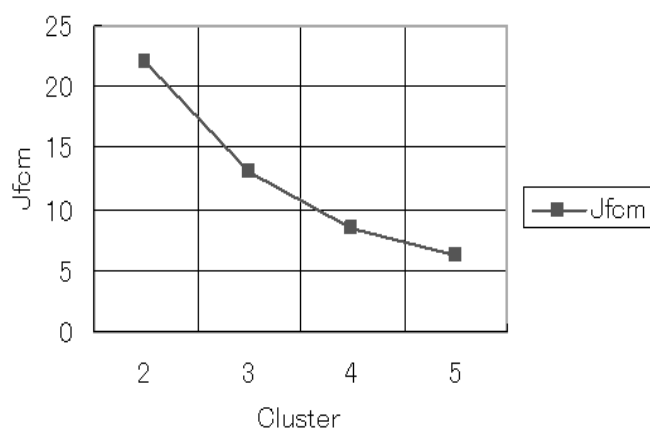


図10 目的関数の増減、変数(女性、子供)

woman-child(基準化)

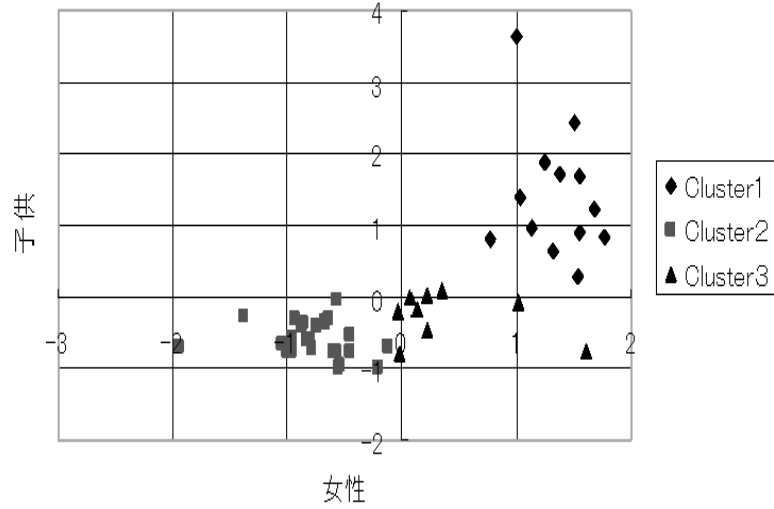


図11 分布図、変数(女性、子供)

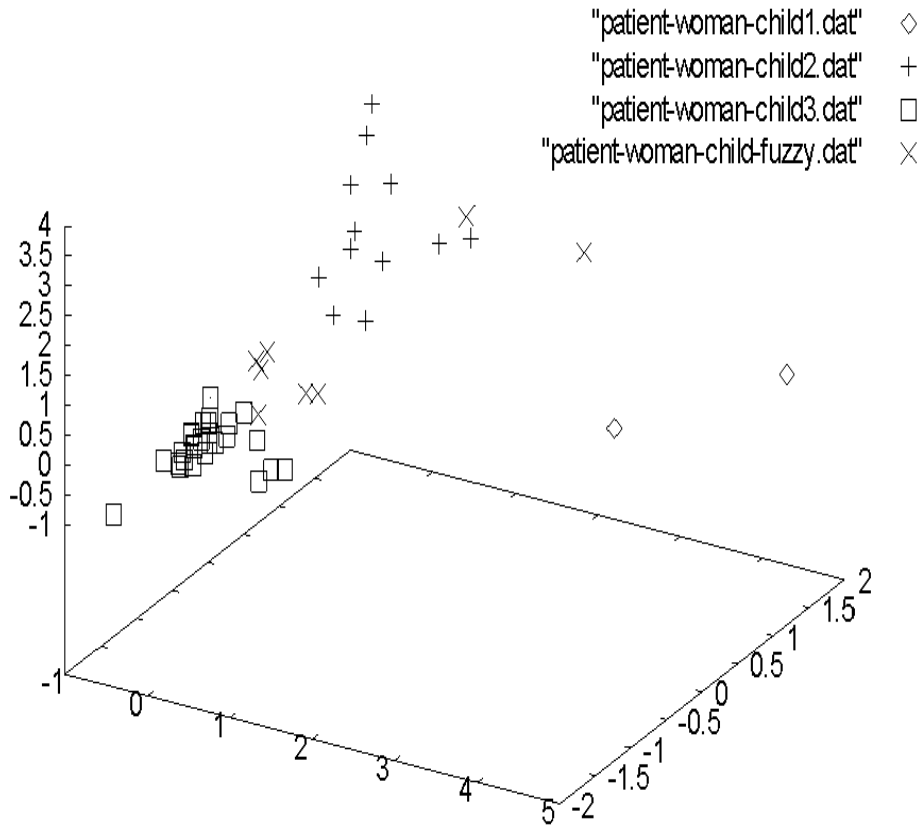


図12 分布図、変数(総疾患数、女性、子供)