

空間ドキュメント管理システムのユーザビリティ向上と実用化に関する研究

研究分担者 有川 正俊 東京大学空間情報科学研究センター准教授

研究要旨

地理空間情報処理を誰でもが簡単に行うためのソフトウェアツールとして、空間ドキュメント管理システム(SDMS)の実用化に向けて、機能の充実と安定化、およびウェブによる一般公開の準備を行った。具体的には、一般デジタルドキュメントに、簡単なタグを埋め込むことにより、簡単に自然に属性情報と位置情報を利用できる枠組みである空間アノテーション機能の精緻と安定化を試みた。また、空間分布の可視化機能を豊かにして、簡単に効果的な空間分布の図を作成できる機能を追加した。さらに、健康危機情報をウェブから集めるためのニュース収集機能をテスト的に実装し、実証実験として、国立保健医療科学院の健康危機管理支援ライブラリーシステム(H-CRISIS)の「健康危機情報に関するマスコミ発表のサービス」のウェブページを対象にその有効性のテストを行い、良い結果を得ることができた。

A. 研究目的

情報伝達技術の発達による社会環境の変化に伴い、健康危機に対して地域保健が対処しなければならない課題が多様化及び複雑化しつつある。しかし、現在の地域保健行政において、現状把握、問題抽出、原因分析等は、社会に存在する情報の一部の情報である保健・医療・福祉に関する統計調査等を画一的な側面から分析しているため、正確な現状把握及び本質的原因の解明を行うことは困難である。この状況を打開し、迅速かつ効率的に現状把握等を実施するためには、地域空間情報を的確に把握し、瞬時に事態を把握することを要するとともに、その原因に関する推論を下しうるソフトウェアの開発が厚生労働行政上、急務である。

申請者らの研究チームでは、上の目的に合う空間ドキュメント管理システム(SDMS)を開発し、その試行版を国立医療科学院の健康危機管理支援情報システムにおいて公開して、保健所などにおいて利用可能な状況にした。しかし、このシステムには以下の課題がある。第一に、現在のSDMSでは地図の操作性が低い。第二に、点的イベントを表示することができるが、線や面な

ど高次元のイベントには十分に対応できない。第三に、集中している地点を簡易に表示するという、健康危機管理において必須と思われる機能を装備していない。これらの問題を解決すべく、現在のSDMSを改良し、日常的にも、健康危機発生時の非常時にも利用できる汎用的なシステムを開発する。

B. 研究方法

平成20年度は、SDMS(図1)の利便性と操作性の向上を中心に研究開発を進めた。平成19年度に、SDMSに導入した空間アノテーション機能は、デジタルドキュメントから抽出されたPOI(Point of Interest)集合に対して、任意の属性値(たとえば、患者数、温度など)を、文章中でジオタグ(ジオタグの例:<SPA attrname1="気温" attrvalue1="25.6">東京大学本郷キャンパス</spa>)を記述するだけで、日常のドキュメント作成および加工業務の一環として、SDMSを用いた空間密度分布を自然に表現できるという点で意味がある。平成19年度の段階では、SDMSの中の空間アノテーション機能は、プロトタイプとしての実験的な導入であり、一般配布する品

質の機能とはなっていなかった。平成20年度は、この空間アノテーション機能を精緻化し、一般配布用のSDMSの基本機能として実装を行った。

SDMSの可視化機能の強化は強く望まれていた。具体的には、平成19年度までのSDMSでは、点密度分布の表示方法は、単に点群をそのまま表示したり、ラベル群として表示するものであった。ただし、同じ位置にPOIが複数個マッピングされた場合は、その多重度が分かる必要があるため、POIの点グラフィックスの上にマウスを重ねることにより、多重度を表示するという対話インタフェースを用意していた。しかしながら、これらの離散的な点密度分布は、補間して連続面として表現する方が視覚的に効果的であり、効率良く正しい情報を伝達できる。このような要求のもと、カーネル(密度推定)法濃淡図と、単純濃淡図の2種類の面的な可視化の手段をSDMSの基本機能として導入した。

ウェブ上の健康危機ニュース情報の自動収集・自動マッピング機能の検討と実装を行った。つまり、SDMSを使って、いろいろな機関から発表されている非定型なニュース情報をインターネットから収集し、ニュース情報の位置を地図として表示する手法を検討し、プロトタイプシステムを開発し、その有効性の検証実験を行った。具体的には、ウォッチするウェブページのURLと自動的に取得する時間と繰り返し周期を設定し、そのウェブページに更新があったかどうかを自動的に確認するとともに、新しいニュースの部分を自動的に抽出し、地図上にPOIとして表示するITサービスの方法の検討・実装・実証実験を行った。今年度は、特に、国立保健医療科学院の「健康危機管理支援ライブラリーシステム(H-CRISIS)」の「健康危機情報に関するマスコミ発表資料のサービス」を対象にこの枠組みの有効性に関して実証実験を行った。

ウェブを通した、最新版のSDMS 3.3 β版の実験的配布方法を検討し、実験的に一般配布を始めた。一般配布を通して、ユーザからの要望を聞き、より安定した、より便利なツールへと改善していくことも目的としている。以前は、健康危機管理支援ライブラリーシステム(H-CRISIS)だけで実験的に公開していたが、

一般から利用したいという要望があり、またツールとして安定性が増したため、東京大学からも実験的に公開することにした。ただし、利用者が自由にダウンロードできるという形式の公開ではなく、利用許可を得てから、一時的なURLとパスワードを通知し、ダウンロードしてもらう形式を取っている。

C. 研究成果

空間アノテーション機能(図2)を使うことにより、デジタルドキュメントに自然に属性値の記述を行うことができ、空間分布の属性値のデータを負担無く準備できる新しい利用者環境を示すことができた。文章の中で属性値を記述する場合、データはアドホックなものが主であり、複雑な構造を持つ属性値あるいは多くの種類の属性値の記述は行われまいだろうという仮定のもと、現在は1つのドキュメントで記述できる属性値の個数を4つと制限して、利便性の面で記述形式を簡単化している。

空間密度分布の簡易的可視化機能で用意した二通りの方法は、空間データの性質により使い分けることを仮定している。カーネル法濃淡図(図3(a))は、位置における多重度の値を正確に面として表現するのに適しているが、計算コストがかかるために表示には時間がかかる。一方、単純濃淡図(図3(b))はそれぞれの点からの減衰率だけで面を構成するために瞬時に表示させることが可能であるが、サンプル点の間の値が不適切になる場合もある。これらの可視化機能は、空間アノテーション機能とも連携しており、ユーザ定義の属性値の可視化機能は実現されている。一方、属性値の性質により、この用意した2つの可視化だけではカバーできないものもあり、今後は、さまざまな性質の連続面の表現が可能になるように、より高度な可視化手法あるいは補間手法を導入する必要があることが分かった。

ウェブ上の健康危機ニュース情報の自動収集・自動マッピング機能(図4、図5)の実現方法を検討し、実際にウェブからニュース情報を自動収集する実証実験を行った。実証実験の対象として、H-CRISISの「健康危機情報に関するマスコミ発表資料のサービス」のウェブページを

選んで、ここに集中して集まってくる健康危機ニュース群の地図化の自動化を試みた。この実証実験を通して得られた知見としては、このウェブページ自体はももとのニュースのソースであるウェブページをURLを使って指しているのもう1段階以上深く、つまり木構造を辿り、より深い段階での情報収集を行う必要があるが、その辿ったURLが指している(各地方組織が発信している)デジタルドキュメントは、その種類も形式も非定型であり一般には自動処理が困難であり、SDMSがこの点からも大変有効であることが改めて分かった。このように、非定型ドキュメントからの自動情報抽出および自動マッピングは便利な機能であることは確認できたが、一方で、正確な情報の共有をしたい場合は、非定型ドキュメントではなく、簡易な定型ドキュメントで配信する方が効率は良いと考えられる。たとえば、1つのドキュメントの中に、地名や住所など複数の間接位置参照情報があり、実際に存在する地理オブジェクトの数よりも多くのPOIが生成する場合が多い。しかしながら、利用者は、不正確な

情報が混ざった冗長なPOI群であっても、ニュースが発生したことが分かることは重要であり、非定型なドキュメント中心のリアルタイム情報収集・通知システムは有効であることは示せた。今後は、これらのニュース情報の公開はRSSなどを用いて行い、ある程度の定型化を進めることにより、自動収集するニュース情報の品質が上がり、また冗長性もなくなると考えられる。

SDMS 3.3β版の実験的配布に関しては、多くのユーザが出てきたときに、SDMSの地図サーバはきちんと動作できるのかというユーザ数のスケーラビリティの問題が懸念されたが、今年度の実験では、地図サーバの性能が低くなるために、ユーザから不満が出るという状況にはならなかった。ユーザからのフィードバックを如何に実現するかに関しては、ウェブサイトやメーリングリストの充実が必要である。このように、一般配布にともないユーザ対応のための人的リソースが必要であるが、現実にはそれを確保するのが困難であるという問題も明らかになった。



図1. SDMS の基本機能である、ジオパース、ジオコード、自動マッピングの一連の流れのイメージ



図 2. 空間アノテーション機能のイメージ

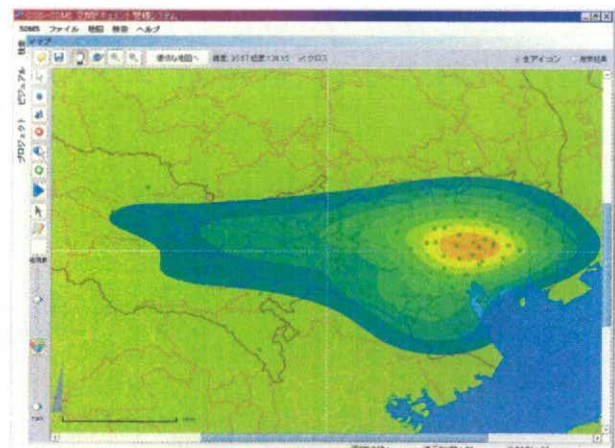


図 3(a). カーネル (密度推定) 法濃淡図の例

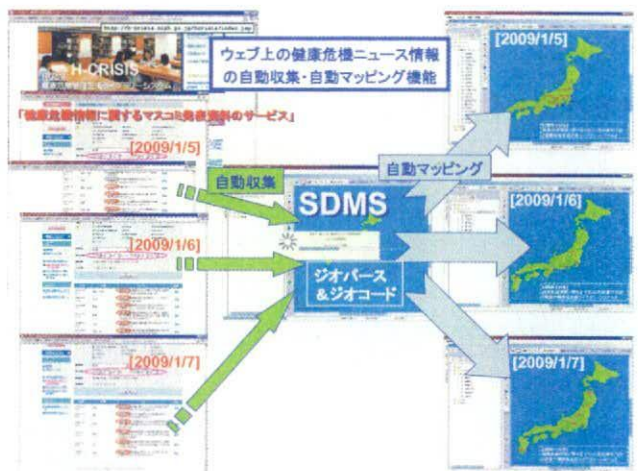


図 4. ウェブ上のニュース自動収集・地図化

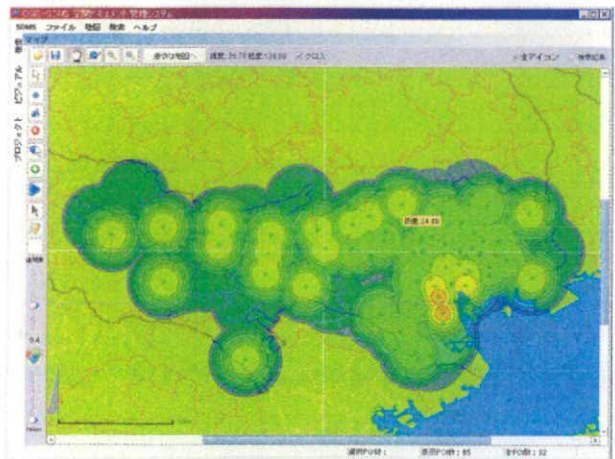


図 3(b). 単純濃淡図の例



図 5. ニュース自動収集の操作パネルの例

D. 考察

SDMS の開発と研究に関する以下の 2 点に関して考察を行う。

(1) 空間アノテーション機能と空間分布の可視化の強化

われわれの本研究の視点は、空間ドキュメント管理システム (SDMS) という新しい枠組みの理論的な体系化だけではなく、実際に保健所などの現場において、感染症 (早期対応) などに有効な実用システムに仕上げ、簡単に、適切に利用できるソフトウェアツール環境を実現することにある。

本年度は、その中でも特に、空間アノテーション機能と空間分布の可視化に力を入れた。従来の SDMS では、一般的な属性を取り扱うことができなかつたので、複雑な空間分布は表現できなかった。そこで、Web 2.0 で多く用いられているアノテーション機能を追加し、一般ユーザが簡単に、一般的な属性値を位置情報とともに、一般デジタルドキュメント上で表現できる機能を実現した。この機能は、まだ実現したばかりであり、実際に利用された事例が少ないので、来年度は、それらを実際に利用してもらうために、ウェブなどを通して、その機能の紹介に力を注ぐ予定である。また、使い勝手の点でも改善の要求が出てくると予想され、それに対応する予定である。

可視化機能は空間アノテーション機能とも連動し、点位置オブジェクトの多重度をうまく可視化するための機能の充実を図った。従来の SDMS では、点やアイコンだけで多重度をグラフィック表示していたが、空間分布の濃淡が直観的に分かりにくいという意見が多くあり、これに対応したものである。この機能は、簡単に空間分布を可視化できるために、利用者からは好意的な意見が聞かれた。しかしながら、温度や雨量など

の空間密度を表現するための正確な可視化機能はまだ提供しておらず、これらの多様な空間分布にも対応できるように、可視化機能を充実して行く予定である。

(2) ウェブからの自動収集とその意義

健康危機事例を分析するには、地理的な分布情報はきわめて重要な情報であるが、それらは一般には入手が困難であるというのが従来状況であった。しかしながら、近年、それらの地理的な分布の情報がウェブ上で公開される兆しが出てきた。

われわれが試験的に開発した、ニュース情報収集機能は、これらの地理的な分布の情報をウェブ上から自動的に収集し、地図上に描画も可能なシステムの開発に係る研究であり、この利用方法の有用性は大変高いと考えられる。一方、このウェブの自動情報収集機能の妥当性について吟味が必要である。つまり、収集した情報と健康危機情報との関連性について考察をして、実用的なシステムであるかどうかの検討が必要であり、今後、実利用を通して、その意義を確認していく予定である。

われわれの基本的考え方は、自動情報収集機能の妥当性に関しては、基となるデータの信頼性の問題であり、これらに関しては、われわれが開発しているソフトウェア SDMS では解決できない問題である。しかし、現在の社会システムの一般的発展方向としては、透明性の確保、情報公開へ進むのは明白であり、今後はウェブに出される情報の信頼性はますます上がると考えられる。また、多様性も増すと考えられるので、ウェブの自動情報収集機能の妥当性・重要性が今後ますます明らかになると予想している。

E. 結論

来年度は、最終年度であり、SDMS の完成度を上げ、実利用に耐えうるものに仕上げる。また、新たに検討・実装する機能として、RSS の

枠組みを用いたウェブ上の健康危機ニュース情報の、自動収集・自動通知・自動マッピングの機能の実現と実証実験を行う。ウェブの世界では、リアルタイムの情報を通知したり、大量の情報から必要な情報を自動的に選択する枠組みとしてRSSが一般化しつつある。健康危機情報の分野においても同様にRSSをベースとする情報公開や情報収集の枠組みが近い将来一般的になるのは確実であり、これを仮定して、RSSによる健康危機ニュース情報を地図上にマッピングする枠組みを試作し、その有効性と問題点などを整理し、来るべきリアルタイム健康危機ニュース情報管理の枠組みの体系化を試みる。さらに、SDMSの普及を考えて、SDMS自身の操作方法および空間情報処理の基礎の教育として、SDMSを使った e-learningの教材を整備し、保健医療分野における空間情報システムの利活用の普及をめざす。

F. 研究発表

1. 論文発表

[1] 浅見泰司, 有川正俊, 白石陽, 相良毅 (2008) 「健康危機管理のための空間ドキュメント管理システム」, 保健医療科学 57 巻 2 号, 特集: 地域診断・症候サーベイランスに向けた空間疫学の新展開, pp. 137-145.

2. 学会発表

[1] Morishige Ota, Masatoshi Arikawa, Hideo Makino, Koichi Kubota, Atsuyuki Okabe (2008) “The Research Initiative on a Curriculum for Geographic Information Technology in Japan”. The 6th ASIA Symposium on Geographic Information System From Computer and Engineering View (ASGIS'2008), Niigata, pp. 5-12.

[2] Hideki Kaji and Masatoshi Arikawa (2008) “Personal Location Based Services on Place-enhanced Blog”. ACM International Conference on Multimedia, Vancouver, Canada.

[3] Ken'ichi Tsuruoka and Masatoshi Arikawa (2008) “User-Generated Audio Tours with

Synchronized Maps on Mobile Media Players”. Asia GIS Conference 2008, Busan, Korea.

[4] Masatoshi Arikawa and Hideki Kaji (2008) “Location-aware Personal Life Content Aggregators on Place-enhanced Blogs”. 8th International Symposium on Web and Wireless Geographical Information Systems (W2GIS), Shanghai, China.

[5] 白石 陽, 有川 正俊, 浅見 泰司 (2008) 「SDMS: 空間ドキュメントの地図化と管理のためのユーザ中心型ツール」. 日本地球惑星科学連合 2008 年大会.

G. 知的財産権の出願・登録状況 特になし

謝辞

本研究を遂行するにあたり、アドレスマッチングのエンジン部分を本システム向けに改良を加えて利用させていただきました相良毅氏に感謝します。また、相良毅氏には、空間ドキュメント管理システムの設計の際に多くの有意義なアドバイスとコメントをいただいた。アドレスマッチング処理の一部では、国土交通省 国土計画局 国土情報整備室が提供している「街区レベル位置参照情報」を利用させていただいている。背景地図は、国土交通省 国土地理院が提供している「数値地図 25000 (空間データ基盤)」を利用させていただいている。ソフトウェア配布に関しては、国立保健医療科学院 健康危機管理支援ライブラリーシステム (H-CRISIS) に支援いただいている。

参考文献

なし

研究成果の刊行に関する一覧表

雑誌

発表者氏名	論文タイトル	発表誌名	巻号	ページ	出版年
郡山一明, 片岡裕介, 竹中ゆかり, 浅見泰司, 高橋邦彦, 丹後俊郎	健康危機管理と小学校 欠席状況サーベイランス	保健医療科学	第57巻第2号	130-136	2008
浅見泰司, 有川正俊, 白石陽, 相良毅	健康危機管理のための 空間ドキュメント管理 システム	保健医療科学	第57巻第2号	137-145	2008

健康危機管理と小学校欠席状況サーベイランス

郡山一明，片岡裕介，竹中ゆかり，浅見泰司，高橋邦彦，丹後俊郎

2008

保健医療科学

(国立保健医療科学院)

特集：地域診断・症候サーベイランスに向けた空間疫学の新展開

健康危機管理と小学校欠席状況サーベイランス

郡山一明¹⁾²⁾, 片岡裕介³⁾, 竹中ゆかり²⁾, 浅見泰司³⁾, 高橋邦彦⁴⁾, 丹後俊郎⁴⁾

¹⁾ 北九州市危機管理参与, ²⁾ 救急救命九州研修所,

³⁾ 東京大学空間情報科学研究センター, ⁴⁾ 国立保健医療科学院技術評価部

Health Crisis Management and Absentee Surveillance in Elementary School

Kazuaki KOHRIYAMA¹⁾²⁾, Yusuke KATAOKA³⁾, Yukari TAKENAKA²⁾,
Yasushi ASAMI³⁾, Kunihiko TAKAHASHI⁴⁾, Toshiro TANGO⁴⁾

¹⁾ Supervisor of Crisis Management in Kitakyushu City, ²⁾ Emergency Life Saving Training Academy of Kyushu,

³⁾ Center for Spatial Information Science, University of Tokyo,

⁴⁾ Department of Technology Assessment and Biostatistics, National Institute of Public Health

抄録

健康危機管理における小学校欠席状況サーベイランスの有用性について解説した。欠席状況をサーベイランスシステムとして使用するためには、日常的に存在する小学校在籍者数及び欠席状況の学校間較差を解消する必要がある。この較差を解消するために標準化欠席比という考え方を導入した。本方法を用いることで、実際に起きた地域の健康危機事態のアウトブレイクを早期に把握することが可能であった。また、GISを用いることでインフルエンザ流行期の欠席状況には早期より空間集積性が見出された。健康危機管理におけるサーベイランスの目的は、危機拡大を抑制するための予防行動強化を起こすきっかけを得ることにある。小学校欠席状況サーベイランスは、健康危機事態の早期把握に有用であり、空間-時間集積性検定を応用することでさらに有用性が高まると考えられる。

キーワード： 健康危機管理, サーベイランス, 欠席率, GIS, 空間-時間集積性

Abstract

The utility of the elementary school absentee surveillance in the health-crisis-management was explained. As a surveillance system, it is necessary to dissolve interscholastic difference of the elementary school enrollment and the absence situation that exists daily to use the absence situation. To dissolve this difference, we introduced the idea of standardized absentee ratio. The outbreak of the health crisis in the region that actually occurred was able to be detected by using this method at the early stage. Moreover, at the early stage, the space cluster was found to the absence situation of the influenza fashion period by using GIS. The purpose of surveillance in a health- crisis-management is to obtain the chance to cause the preventive action to control the expansion of crisis. We consider that the elementary school absentee surveillance is useful for detect at the early stage of the health-crisis-situation. Moreover, we consider that the utility of surveillance rises further because it applies spatial-time scan statistic.

Keywords: health crisis management, surveillance, absentee ratio, GIS, spatial-time clustering

〒807-0874 福岡県北九州市八幡西区大浦3-8-1

救急救命九州研修所

3-8-1 Oura, Yahata-nishi-ku, Kitakyushu, Fukuoka, 807-0874 Japan.

Email: kooriyama@fasd.or.jp

1 はじめに

危機は自然と文明の共同産物である。特定地域の危機にすぎなかった風土病が物流や交通路の発達によって全世界の日常的な危機となったように、人間生活の発展・文明化は絶えず新たな危機を生み出していく。また、都市化、すなわち生活空間の密集化は単一危機で大量被災者を生む上に、都市システムはそれぞれの機能が関連して構築されているがゆえに、次々とシステムを破綻して被災の場を増やし、複合危機を形成していくという構造的な脆弱性を持っている。このような現代社会では危機管理、とりわけ健康危機の兆候を把握するサーベイランスの重要性は益々大きくなってきている。

現在実施されている感染症をはじめとする健康危機に関するサーベイランスの多くは、予め原因リストを作成しておき、その数の変化を感知する方法が用いられている。この方法では原因判明に至るまでに検査を含めて一定の時間がかかることは避けられず、原因が判明した時には既に危機が拡大していることが少なくない。そもそも新興感染症をはじめとして、原因リストが準備されていない危機には対応できない。危機のアウトブレイクという用語は曖昧に使われがちであるが、その真意は次の2つの場合に集約できると考えられる。すなわち、危機の要因が地域内で、①一定の量を超えた場合、②空間的拡大をしていく場合、である。この観点にたてば、現在、わが国で実施されている健康危機に関するサーベイランスシステムは迅速性に乏しく、また数の増減は空間的拡大を反映しないので、危機のアウトブレイク把握には不十分である。

健康危機管理におけるサーベイランスの目的は、危機の拡大を抑制するための予防行動強化を起こすきっかけを得ることにある。求められる最初の条件は、原因の正確性ではなく危機兆候把握のスピードであり、そのためには人間の社会生活の中から健康危機に関係する現象の突然の変化を検知する prospective なシステムを構築することが望まれる。具体的には定まった地点で日常的に観察されている現象の中から、危機に関連した変化を把握するサーベイランスシステムである。

小学校は地域毎に相当数が存在し、毎日一定数の児童が継続的に登校してくる。そして在籍児童が健康被害に陥った場合には、その原因が感染症であれ化学物質であれ、欠席という形で反映される。

我々は小学校のこの特性を健康危機兆候把握に応用できないかと考え調査・研究を進めてきた¹⁾。

本稿ではこれまでの成果の一部を示し、健康危機管理システムとしての小学校欠席状況サーベイランスの可能性について解説する。

2 小学校欠席状況サーベイランスに期待される特徴とそのための準備

小学校は校区によって地域性が限定されているうえに在

籍している年齢層と数が一定である。すなわち、欠席状況の解析にあたっては、年齢調製の必要がなく、在籍者数を分母として欠席率を算出することで健康危機に陥っている状況（危機の大きさ）を異なる小学校間で数学的に比較できるという大きな利点を持っている。このことは、空間疫学の方法²⁾を用いることで異常地域の集積性を検出でき、経時的に解析することで空間的拡大も検出できるという極めて魅力的な可能性にも繋がっている。

欠席率を学校間比較に使用するためには、平常時の欠席率はどの程度なのか、在籍者数による影響はないのか、平常時の欠席状況に学校間較差はないのか、欠席率の異常をどのように定義するのか、などについて検討しなければならない。そこで、我々はまずこれらについて北九州市を対象として、市内の公立小学校のデータを収集して検討した。北九州市は九州北部に位置する面積485.55Km²、人口1,000,559名（平成16年6月1日）の政令指定都市である。調査時点での市立小学校総数は132校、総児童数は52,339名であった。

132小学校の在籍者数分布を示す。301名から400名の小学校が最も多かったが、全体の36%は300名以下であり100名以下の学校も8校存在した（図1）。

このような在籍者数分布の元で、市内で健康危機が発生していなかった15週間の各小学校の欠席率を定期的に週1回調べた。132校の欠席率は2.0±1.2%であった。縦軸に各週のそれぞれの学校の欠席率、横軸に小学校在籍者数をグラフにプロットすると、在籍者数が少ない小学校において欠席率のばらつきが大きくなっていることが分かった（図2下段）。これは在籍者数が少ない小学校では欠席者数の増減の影響が大きく反映されて不安定になるからと考えられる。さらに規模が等しい小学校においても欠席率に大きな較差がある場合があることが分かった（図2上段）。

この結果から、各小学校の欠席状況を比較するためには、欠席率を単純に比較するのではなく何らかの標準化を図る必要があることが示された。

標準化の方法として、各学校別に健康危機事態が起きていない期間を基準として、この間の平均欠席率を期待欠席率（E）とし、実際に観測された欠席率を（O）として、O/E比をとったものを標準化欠席比とした。当然のことながら標準化欠席比を用いることで学校間較差はなくなり（図3上段）、在籍者数によるばらつきも小さくできた（図3下段）。

健康危機が発生していない時期の学校の標準化欠席比が m (平均) + 2σ (標準偏差) を超えた割合は0.03であったので、 $m + 2\sigma$ をカットオフ値としてカットオフ値以上の場合を「異常」とであると定義した。すなわち、カットオフ値は $p < 0.03$ に相当する。

3 実際の事例による検証

本方法の有用性を過去の健康危機事態を用いて検証した。用いたのは1996年に埼玉県の越生市で起きた集団下

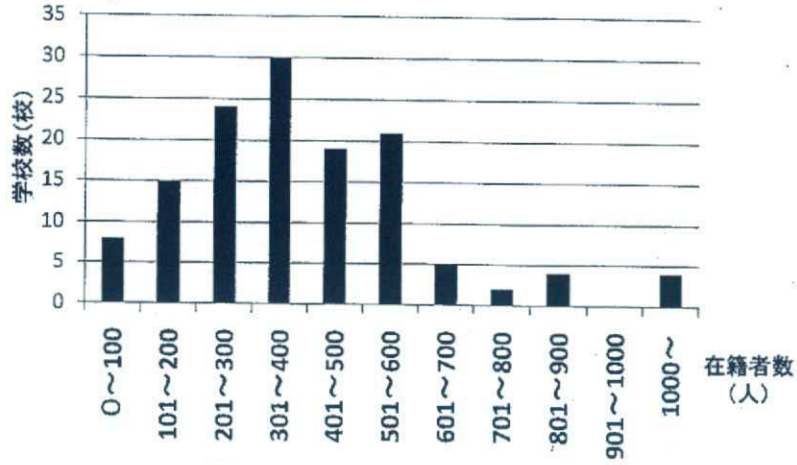


図1 北九州市小学校の在籍者数別学校数

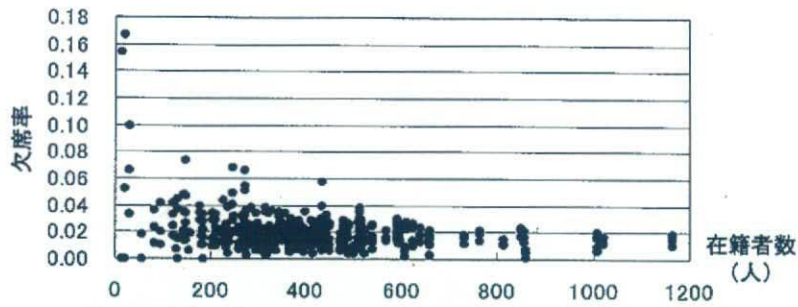
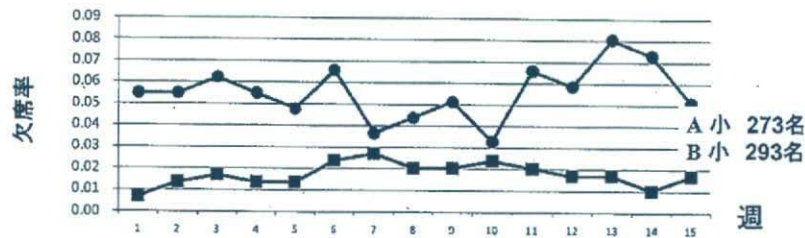


図2 欠席率:学校による欠席率較差(上段), 在籍者数の関係(下段)

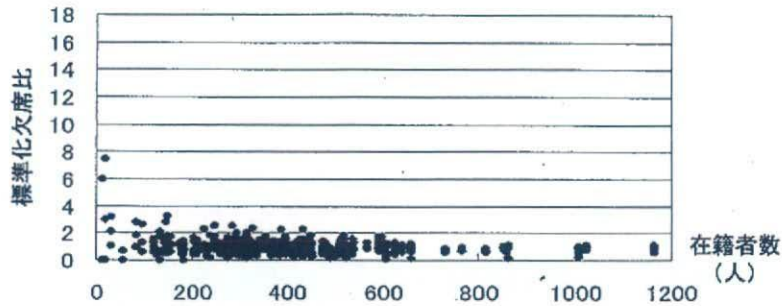
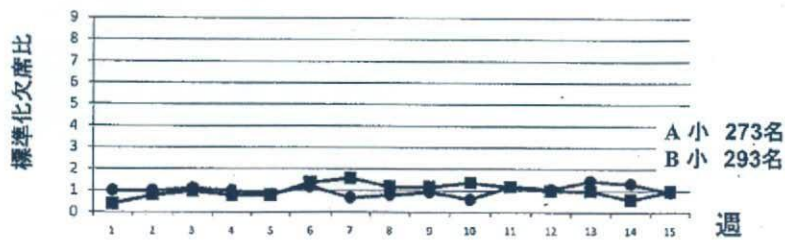


図3 標準化欠席比:学校間較差(上段), 在籍者数との関係(下段)

病症である。

[事態概要]³⁾⁴⁾

1996年(平成8年)6月10日, 埼玉県越生町で下痢・腹痛等の症状で小中学校の児童生徒が多数欠席しているとの一報が入り, 同日から調査したところ町内の住民に同様の患者が多数いることが分かった。6月17日, 有症者の検便からクリプトスポリジウム(原虫)が検出され, その後, 同町の水源, 水道水からも検出されたため, 水道水を介したクリプトスポリジウムによる集団下痢症であることが判明した。

[標準化欠席比による解析]

埼玉県越生町と隣接する毛呂山町, 鳩山町に所在する10小学校の欠席状況について, 6月3日から6日までの4日間を基準として17日までの標準化欠席比を10小学校全体と学校別に算出した(図4)。10小学校全体の標準化欠席比は6月10日から異常となった。学校別にそれぞれの

標準化欠席比とカットオフ値を検討したところ, 越生町2小学校が6月7日の時点から異常となり, その後も継続して上昇していた。隣接する毛呂山町の毛呂山小は標準化欠席比の大きさは小さいものの10日から異常を示し, そのピークは越生町の2小学校の欠席状況のピークと同様であった(図5)。

[解説]

標準化欠席比を用いることで, 越生町周辺に何らかの健康危機事態が起きていることを早期に把握できた。本方法では毛呂山小でも欠席状況の異常が観察された。報告書には越生町以外での健康危機についての記載はないが, 毛呂山町の水源になっている阿諏訪川でもクリプトスポリジウムが確認されたことを記載しており, 毛呂山小の欠席状況のピークが越生町の欠席状況と一致することを考えれば毛呂山町でも少ないながら健康危機を来していた可能性も否定できない。

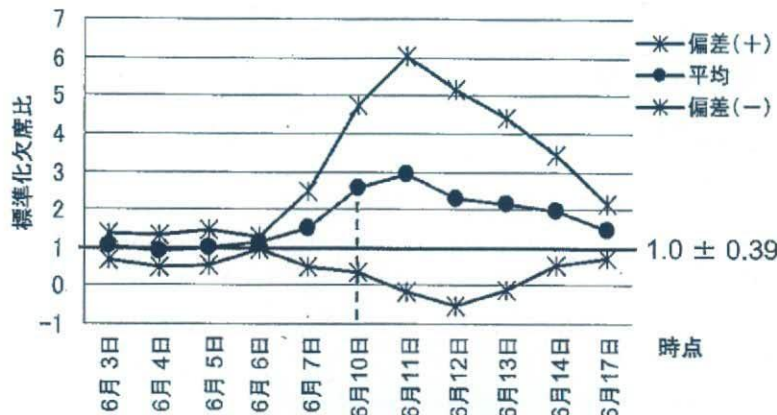


図4 越生町近隣の10小学校の標準化欠席比の推移

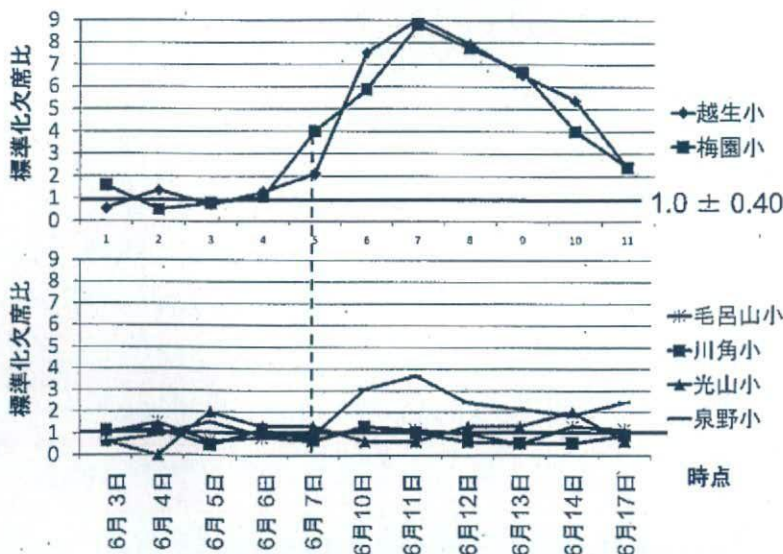


図5 標準化欠席比の推移:越生町(上段), 毛呂山町(下段)

小学校欠席状況サーベイランスは健康危機兆候把握に有効であった。

4 GISを用いた地域感染症流行への応用

2005年の北九州市において市立の132小学校の欠席状況を1月11日から3月22日まで定期的に週1回調べた。この年度の市内全体の平常時標準化欠席比は 1.0 ± 0.57 であった。市内全体の標準化欠席比の変化を示す(図6)。

定点観測ではこの時期に地域で流行していたのはインフルエンザのみであり、小学校の欠席状況はインフルエンザの流行によるものが殆どと考えられた。実際、インフルエンザ定点値と標準化欠席比の相関係数は0.85であり、両者には強い相関が認められた(図7)。市内全体の標準化欠席比がカットオフ値を超えて異常となったのは1月25日であった。定点観測値がインフルエンザの注意報基準値である10を超えたのは2月1日であった。

学校別にみると1月11日の時点で標準化欠席比が自校のカットオフ値を超えたものが13校あり、18日の時点で超えたものが50校と経時的に増加していった。

それぞれの学校で観察された標準化欠席比をもとに、クリギング法によって空間補間を行い、GISを用いて地図上に視覚化した。1月11日の時点で欠席状況が異常な地域に集積性があることが示され(図8)、18日には一気に拡大していた(図9)。

[解説]

クリギング法は元来、任意の地点のボーリング調査により鉱産物の埋蔵量を過大評価せずに推計するために開発された代表的な空間補間法のひとつである。この方法では座標上の事象を連続的な局地的な変数とみなしたうえで、局地的な空間変動は①全体的傾向、②空間的に相関しているローカルな残差、③ランダムなノイズ、の3つの成分から構成されるとする⁹⁾。この考え方は現在、気象学、農学、

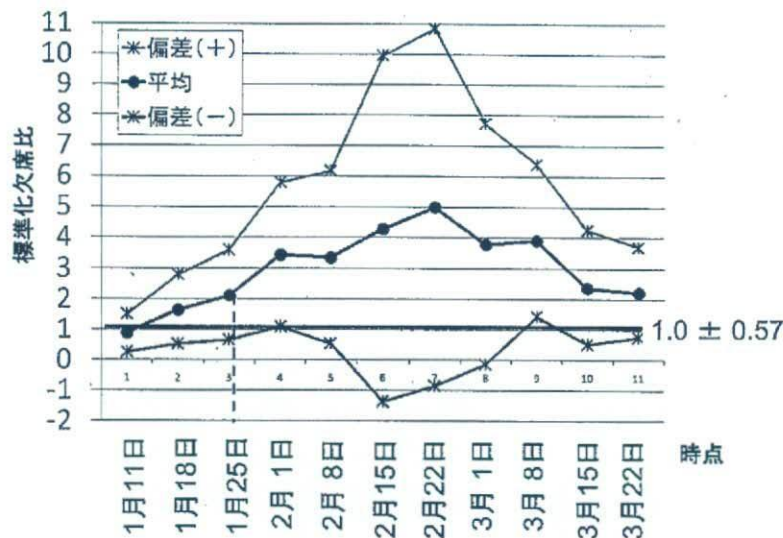


図6 2005年インフルエンザ流行期の北九州公立小学校標準化欠席比の推移

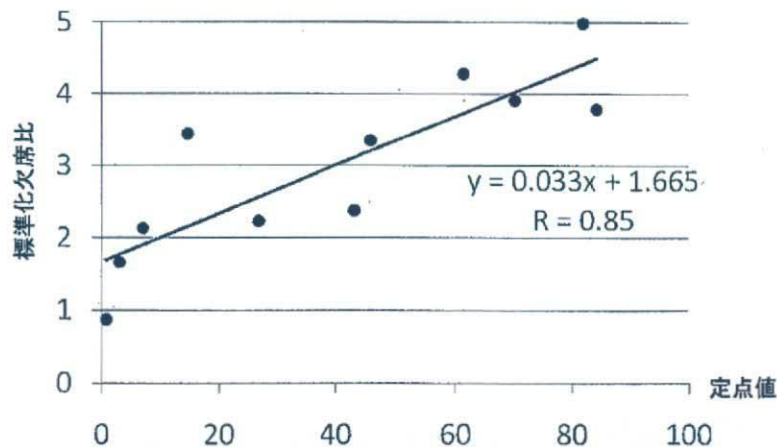


図7 北九州市のインフルエンザ定点値と小学校標準化欠席比の相関

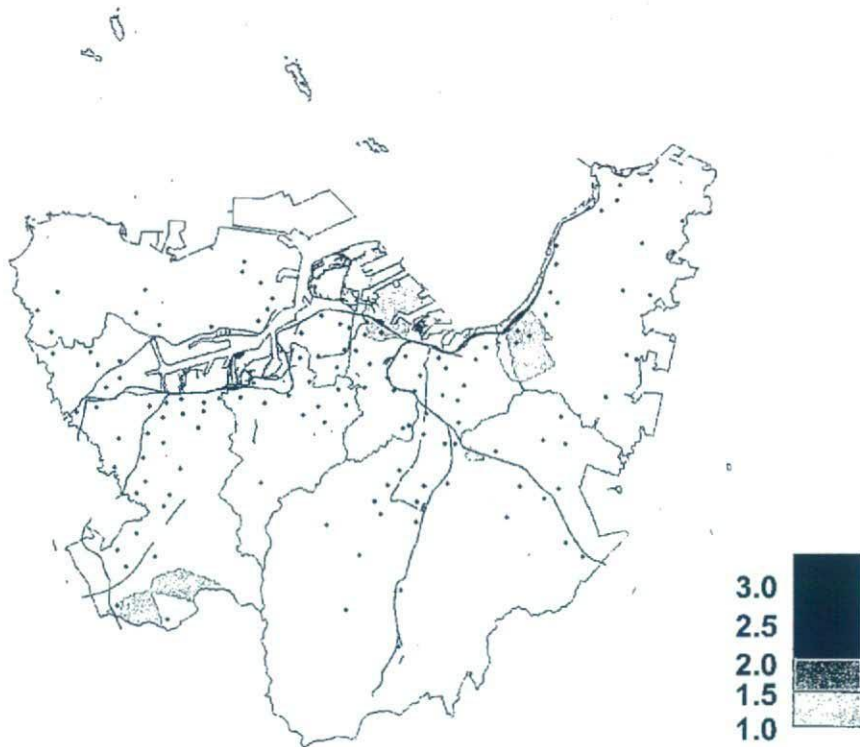


図8 カットオフ値による異常欠席状況小学校の空間集積：1月11日

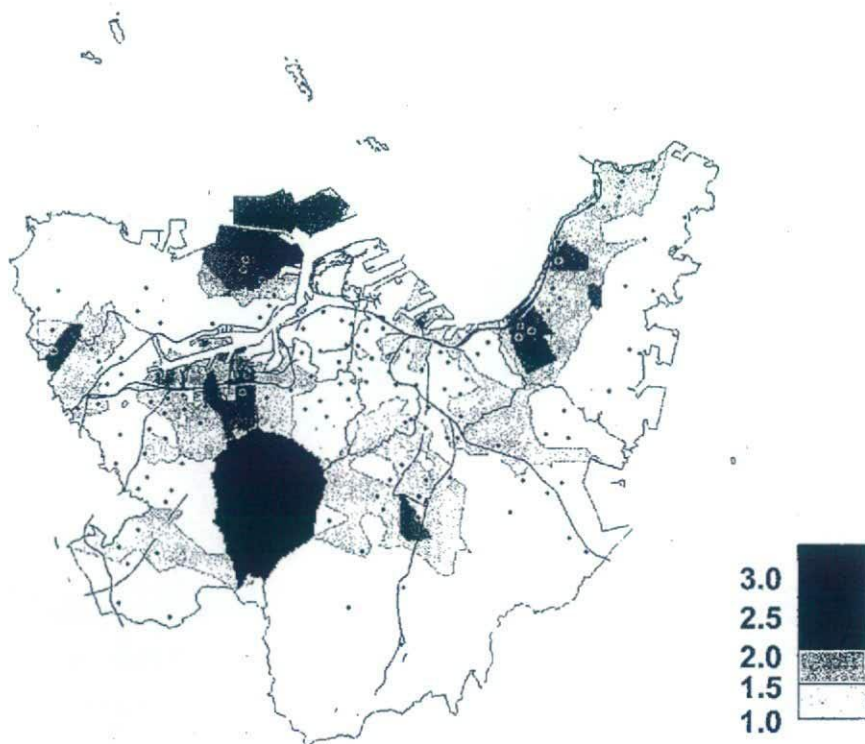


図9 カットオフ値による異常欠席状況小学校の空間集積：1月18日

森林学等に広く応用されている。同様に、地域の欠席状況について空間的な多寡を過大評価せずに評価できるであろう。GISと併用すれば欠席状況の空間的集積性を視覚化して表示できる。

一般に欠席状況に空間的集積性があることは、集積地域の学童が同一の曝露源に接したか、ヒト-ヒト伝播を来たす感染症のアウトブレイクが起きているかの2つの場合に限られる。図8のように異常の集積性が散在して認められる場合はヒト-ヒト伝播を来たす感染症のアウトブレイクが強く疑われる。この結果はインフルエンザの地域発生と拡大状況を反映したものと考えて矛盾はない。この場合、1月11日の時点でインフルエンザのアウトブレイクの可能性を指摘することが可能である。さらにこの結果で注目すべきは、わずか1週間での欠席異常状況の拡大であろう。健康危機にとって兆候把握のサーベイランスの迅速性がいかに重要であるか、その必要性を如実に示している。

これらの結果から、小学校欠席状況サーベイランスは感染症の地域アウトブレイク把握に有効であり、GISを応用することによりさらに有利性を発展できると考えられる。

5 今後の展開とそのための検討課題

我々はこれまでの研究を通じて、小学校欠席状況サーベイランスは一定の工夫を加えれば、健康危機管理のためのサーベイランスとして非常に有用かつ実用的なものになりえると考えている。過去の事例でも危機発生初期に当該地域の小学校欠席率が上昇していたものも少なくない^{6~8)}。しかしながらその殆どは「事態を把握した後に振り返ってみれば初期に小学校欠席率が上昇していた」というretrospectiveな解析にとどまっている。

一方、学校の欠席状況を危機把握のためにprospectiveに収集しようという試みもある。アメリカNorth Carolina州Mecklenburgでは、社会における疾患アウトブレイクの早期発見、バイオテロリズムの早期把握を目的として地域の救急医療機関や中毒センターの患者来院状況とともに、188の公立学校の欠席率を収集するMedical On-Line Surveillance Tool (MOST) というシステムを構築している⁹⁾。ある地点における事象が、その周辺地区の事象を受けて互いに相関関係が発生していないかについては、クリギングをはじめとする空間的自己相関の分析手法があるが、MOSTでも単に欠席率を表示しているのみで、その欠席状況をふまえた考察方法については未だ確立されていない。

欠席状況のような時間的データを観察しつつ空間的集積性を検出するには、現時点でも続いている集積性をいち早く「生きているクラスター」として同定するprospective

な解析方法が必要となる¹⁰⁾。これについては、近年、集積期間と地域について集積性の有無を検定するflexible spatial scan等の方法¹¹⁾¹²⁾が開発されており、欠席状況サーベイランスへの活用が期待される。

欠席状況サーベイランスで最も難しいのは「異常」を定義するカットオフ値の決め方であろう。カットオフ値を高く設定すればspecificityは上がるだろうが危機の兆候把握は遅れるに違いない。低すぎれば兆候把握は早くなるが、false positiveも多くなる。

今後はカットオフ値と欠席者の症候の集積性とあわせて検討し、さらにGISと併用することで、欠席状況をふまえた考察方法を開発し、本システムの実用化に向けて研究を進めていく予定である。

- 1) 浅見泰司. 地理及び社会状況を加味した地域分析法の開発に関する研究 平成16年度~18年度 厚生労働科学研究総合研究報告書.
- 2) 丹後俊郎, 横山徹爾, 高橋邦彦. 空間疫学への招待. 東京: 朝倉書店; 2007.
- 3) 埼玉県衛生部. クリプトスポリジウムによる集団下痢症-越生町集団下痢症発生事件: 1997.
- 4) 保坂三継. クリプトスポリジウムとその水系感染症水情報 1996; 16(3): 8-11.
- 5) 張 長春. 空間データ分析. 東京: 古今書店; 2001.
- 6) 堺市学童集団下痢症対策本部編: 堺市学童集団下痢症報告書. 1997.
- 7) 横浜市衛生局感染症・難病対策課. 市立文庫小学校での下痢・嘔吐者の集団発生について (2005/05/26) 記者発表資料 2005.
- 8) 国立感染症研究所. 小学校における風疹集団発生の調査-岡山市. 病原微生物検出情報 (IASR) 2006; 24: 58-59.
- 9) <http://www.co.mecklenburg.nc.us/Departments/Health+Department/Top+News/Pandemic+Flu.htm>
- 10) Kulldorf M. Prospective time-periodic geographical disease surveillance using a scan statistic. Journal of the Royal Statistical Society, Series A, 2001; 164: 61-72.
- 11) Tango T and Takahashi K. A flexibly shaped spatial scan statistic for detecting clusters. International Journal of Health Geographics, 2005; 4: 11.
- 12) Takahashi K, Kulldorf M, Tango T and Yih K. A flexible space-time scan statistic for disease outbreak detection and monitoring. Advances in Disease Surveillance. 2007; 2: 70.

健康危機管理のための空間ドキュメント管理システム

浅見泰司, 有川正俊, 白石陽, 相良毅

2008

保健医療科学

(国立保健医療科学院)

特集：地域診断・症候サーベイランスに向けた空間疫学の新展開

健康危機管理のための空間ドキュメント管理システム

浅見泰司¹⁾, 有川正俊¹⁾, 白石陽¹⁾, 相良毅²⁾

¹⁾ 東京大学 空間情報科学研究センター, ²⁾ 東京大学 生産技術研究所

Spatial Document Management System (SDMS) for Health Risk Management

Yasushi ASAMI¹⁾, Masatoshi ARIKAWA¹⁾, Yoh SHIRAISHI¹⁾, Takeshi SAGARA²⁾

¹⁾ Center for Spatial Information Science, The University of Tokyo,

²⁾ Institute of Industrial Science, The University of Tokyo

抄録

健康危機が発生した場合には、どこで何が起きているかを的確に把握し、効率的な対応方法を検討することが重要である。ところが、このような場所に関する情報は、これまであまり有効には利用されてこなかった。その原因としては、場所情報自体のIT化の遅れがあげられる。場所情報を扱うシステムとしては、地理情報システム (Geographic Information Systems, GIS) と総称されるアプリケーションがあるが、一般に高度な数理的解析を行うために用いられるものであり、操作にあたっては専門的な知識や技能が必要とされることから、誰でもが扱えるようなシステムにはなっていない。また、場所情報の表現には多くのバリエーションがあり、GISで利用できるフォーマットに加工するためにも経験や知識を要するため、健康危機発生などの緊急時に必要な場所情報がGISに容易に利用できるような書式で届けられることを期待することはできない。場所情報は多くの場合、人間が分かるアナログ形式で、つまり住所や地名や経路記述などの自然言語で記録されている。そのため、データ形式を変換したり、高度な操作技術が必要になるなどの理由で、これらの貴重な場所情報は有効に利用されずじまいになってしまうことがほとんどである。

この問題点を克服するために、従来のGISとは異なる新しい枠組みとして、人間中心情報デザインの観点から、空間ドキュメント管理システム (Spatial Document Management System, SDMS) を基本設計し、ソフトウェアツールとして実現した。そして、その有効性を実利用を通して検証してきた。SDMSは、一般ユーザがパソコン上で日常的に利用しているデジタルドキュメント (電子メール、テキスト、ウェブページ、ワード・ドキュメント、エクセル・ドキュメント、PDFドキュメントなど) をアプリケーションウィンドウにドラッグ&ドロップするだけで自動的に住所や地名を抽出し、経緯度を算出して、地図上で空間分布を提示できるソフトウェア環境を提供する。一般に、従来のGISが専門家向けの枠組みであったのに対し、SDMSは専門知識無しに誰でもがデジタル空間情報を日常的に取り扱える新しい空間情報利用環境であり、本論文ではSDMSの提案の背景と意義、その基本原理を紹介し、今後の展望について述べる。

キーワード： 非定型データ、ジオコーディング、ジオパーシング、空間ドキュメント、人間中心情報デザイン

Abstract

Information concerning "where" is essential to effectively tackle the outbreak of health hazards. However, only limited use of such information has been practiced till now. Slow development of information technology related to spatial information may be accused to this fact. Geographic information system (GIS) is the software specially designed to deal with spatial information for advanced spatial analyses, which unfortunately shuts novice users away due to heavy needs of skills for its operation. Moreover, so much variation of expression for spatial information is available, and its transformation to GIS-ready format needs professional experience and knowledge. Therefore, in the event of health hazard emergency, well formatted data is rarely obtainable which is ready for GIS use. In many cases, the collected data is described in natural language, such as

〒277-8568 千葉県柏市柏の葉5-1-5

5-1-5 Kashiwanoha Kashiwa, Chiba, 277-8568 Japan.

addresses, area names and route descriptions. The necessity of data conversion and special skills to use GIS often discourages people effectively utilizing spatial information.

To remedy this situation, we developed new software called spatial document management system (SDMS), which is designed as a human centered tool for information processing in a framework different from traditional GIS. The SDMS has been validated through its actual usage. SDMS attains automatic process of extraction of addresses and area names, calculation of latitude and longitude, and visualization of their spatial distribution on a map by only drag and drop operation of digital documents (email, text, web page, MS-word document, MS-excel document, PDF document, etc.) that are handled on a PC on a daily basis. Compared to GIS which is designed for experts, SDMS is designed for ordinary people without expert knowledge. This article is devoted to exposition of background and its significance of development of SDMS, its basic principle of operation, and prospect of its future development.

Keywords: unstructured data, geocoding, geoparsing, spatial document, human-centered information design

1. はじめに

臨界事故に見られる原子力ハザード対策、パンデミック被害やテロなどに起因するバイオ・ハザード、事故や犯罪による化学薬品ハザードなど、現代社会において健康危機リスクは多様になってきており、かつ、素早い対応をしないとその被害は大きくなりかねない。このため、健康危機イベントが発生した際に、それへの素早い対応体制整備は厚生労働行政上、喫緊に行わなければならない重要な課題となっている。

事実、このような事態に対応するために、危機発生時に使える大規模システム構築の開発などが行われてきた。しかし、緊急時にしか使われないシステムは、往々にして、ユーザがその操作法を未習熟であるケースが多く、十分に活用できない事が多い。そのため、平常時にもよく使われ、緊急時にも利用できるようなシステムが、実は危機管理上、有効であると言える。

緊急時には、ただでさえ情報伝達が難しい状況であることが多く、情報発信者には受け手側の便宜をはかって書式を整えるなどといった余裕はないと考えた方がよい。このため、定型ではなく、様々な非定型ドキュメントを有効に活用できるシステムが望まれていると言えよう。

本論で紹介する空間ドキュメント管理システム (SDMS: Spatial Document Management System)^{1,2,3,4)} は、上記の目的を具現化するために開発されてきたものである。空間ドキュメント管理システムは、テキストドキュメントに分解しうる非定型ドキュメント (たとえば、電子メール、テキスト、ウェブページ、ワード・ドキュメント、エクセル・ドキュメント、PDF ドキュメントなど) をドラッグ&ドロップという簡単な操作だけで、地図化し、かつ GIS (地理情報システム) などに取り込めるようなファイルに変換するシステムである。自由な書式でありながら、汎用性に優れたデータ書式で空間情報を流通できるシステムが確立できるため、健康危機情報の収集から空間データ化するまでの時間や手間を大きく省くことができる。健康危機の初動時に被害場所分布を速やかに一覧したい際に有効である

が、実際には、地図化の作業は通常業務でも多々必要とされることから、平常時でも頻繁に使われるはずの汎用的システムとなっている。

すでに、SDMS の試行版を国立保健医療科学院の健康危機管理支援ライブラリーシステム (H-CRISIS) で公開し、保健所などで利用可能な状況になっている。ただ、現在の SDMS では地図の操作性が低い、線や面など高次元のイベントには対応できない、集中している地点を簡易に表示する機能がないなどの課題がある。そこで、現在の SDMS をさらに改良するべく開発を行っている。これにより、空間データ化された情報をそのまま一次解析することができ、健康被害症候の特異性を早期に検出し、またその原因や健康被害の空間的拡散の特色を適切に推測する支援システムを構築でき、リアルタイムでの知的地域診断システムの構築およびその行政的利用が可能となる。さらに、ユーザに対しては各種時空間情報の GIS への統合に伴い現状把握、情報の相互関係、問題点把握等に関してより理解しやすい形での表示を可能とする等により、健康で安全な社会の構築の礎となる。このシステムの応用により、迅速な健康被害に対する警告発信、原因究明の初動期支援、平常時における健康関連情報の発信などに寄与するものと思われる。

2. SDMS の概要と基本原理

2.1. SDMS 開発の背景と意義

パソコンを用いた日常的な業務で一般的に扱うドキュメント、たとえば電子メール、テキスト、ウェブページ、ワード・ドキュメント、エクセル・ドキュメント、PDF ドキュメントなどには、住所や地名などの多くの場所情報が記載されている。これらの場所情報は人間にとって可読性・伝達性が高く、頻繁に利用されている。一方、従来の GIS では、経緯度のような座標値といった、機械に扱いやすい数値による場所情報を主対象としてきた。その結果、多くの場所情報を含む一般ドキュメントは、従来の GIS では扱うことができない。

座標のように機械に向く場所情報記述を直接位置参照情

報と呼び、一方、住所や地名のように、人間にやさしい場所情報記述を間接位置参照情報と呼ぶ。住所を経緯度に変換するように、間接位置参照情報を直接位置参照情報へ変換する処理をジオコーディングと呼ぶ。われわれの研究では、ジオコーディングにより、一般ドキュメントをパソコン上で簡単に地図と連携させて扱える枠組みとして空間ドキュメント管理システム (SDMS: Spatial Document Management System) を提案・実現した。専門家のツールである GIS と違い、医療従事者を含む一般ユーザが SDMS を用いることにより、日常的に利用しているドキュメントを空間的に管理できるようになる。

たとえば、重要な会議のお知らせを電子メールで受け取った場合を考えてみる。会場は今まで行ったことがない場所で、メールには会場の住所と名称しか書かれていないとする。このような場合、PC を使いこなしているユーザであれば、メール文中の住所の部分をクリックボードにコピーして、インターネットの地図検索サイトや PC にインストールしてある地図ソフトウェアなどにその住所をペーストし、会場周辺の地図を表示させることができる。SDMS は、住所を抜き出して該当する座標に変換し、地図を表示するという一連の処理を自動で行うため、ユーザはただメールを SDMS のウィンドウにドラッグ&ドロップするだけでよい。

2.2. SDMS のコンセプト

パソコンで扱われているデータは、人間が作成し、人間が読むドキュメントが中心である。そのようなドキュメントの作成と管理を行う環境がパソコンと言ってもよいだろう。一方、現在パソコンで地図を扱う上で一般的な地図ソフトウェアや専門家が利用する GIS は、地図を中心にデータを管理しようとするため、ユーザ自身がドキュメントに含まれる場所情報を抽出して専用のデータを作成し、その後のデータの追加や編集なども全て地図の上で行わなければ

ならない。ほとんどのパソコンユーザにとって、地図は見たいときに見られればよいものであって、わざわざ手間をかけて専用のデータを作成し、維持管理するほどの必要は感じないだろう。

本論文で提案する SDMS は、一般ユーザが日常的に作成・利用するドキュメントを中心に考え、上述のように簡単に自然に地図と連携させることにより、ワープロ・電子メール・ウェブブラウザを使う感覚で、日常的に空間情報を利用・作成する環境を広く提供する。これが Spatial Document Management System, 空間ドキュメント管理システムという名前の所以である。

たとえば、Windows のエクスプローラなど、ドキュメントを整理するためのファイルマネージャと呼ばれるソフトウェアでは、五十音順や時間順による表示機能は不可欠であり、これに基づいてユーザはドキュメント管理を行っている。同様に、ドキュメントを「空間順」で表示・管理するソフトウェアが SDMS の主概念であり、SDMS を用いれば文章中の地理的な内容やドキュメントを作成した場所を表現する位置印 (location stamp, 時刻印 time stamp に対する概念) に基づいて空間順に表示できる。もともと、空間情報は特別なものではなく、われわれ人間は活動の記録を残したり、活動の予定を立てる場合には、必ず地球上の場所に関する情報を使ってコミュニケーションを行っており、むしろ場所情報記述が使えないと日常活動に支障をきたす。この状況は、IT が出現する以前から変わりはないが、古典的な IT では、人間同士がコミュニケーションで使う空間情報記述を取り扱う能力がなかった。SDMS はそれを可能として、人間が日常的に取り扱っているさまざまなドキュメントを地図上で利用できるようにする環境を実現した点で意義がある (図 1)。

また、広義のドキュメントである画像やムービーなどのマルチメディアコンテンツも空間順で管理する重要性は認められている。たとえば、写真ファイルに関しては、GPS

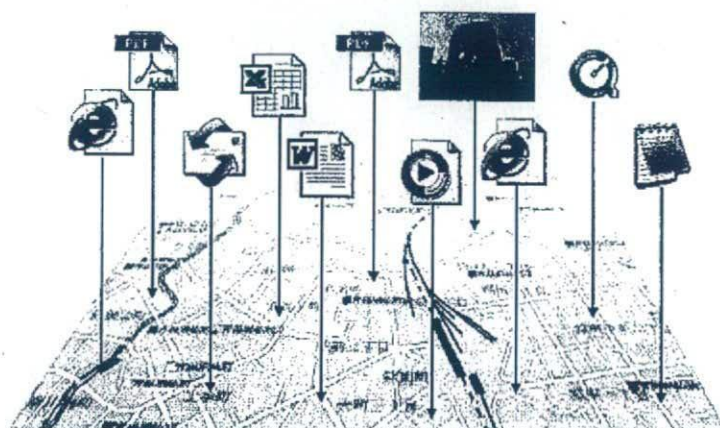


図 1. 空間ドキュメント管理システム (SDMS) のイメージ

日々利用している一般ドキュメント (MS ワード、テキスト、メール、ウェブ、MS エクセル、PDF など) を簡単に地図上の位置に結び付ける利用環境を実現する。ファイルをアルファベット順、時間順で見るがごとく、空間順で見ることが出来る利用環境を実現する。将来は、この機能はデスクトップへ統合されると考えられる。また、EXIF や MPEG7 などの国際標準により、画像やムービーへの位置記述は一般化してきている。

から取得した位置情報を EXIF 形式でメタデータとして埋め込む仕組みはほとんど全ての GPS 付携帯電話で一般的になっている。動画に関しても、ロケの位置情報をメタデータとして記述する国際標準 (MPEG7 など) もすでにある。

2.3. SDMS の使用例

SDMS の利用方法は極めて簡単で、一般ドキュメントをデスクトップ上で、SDMS の地図ウィンドウへドラッグ&ドロップするだけである。ここでは2つの使用例の表示結果を示す。

1つめの例は、東京大気公害裁判原告の住所リストが記載されたエクセルのドキュメントを SDMS で表示した結果である (図2)。この図から本郷通りや早稲田通り、首都高速4号線と5号線の沿線に居住している原告が多いことが分かるので、まずこれらの道路から調査を始めればよいと考えられる。このように健康危機が発生した場合、その要因が地理的なものかそれ以外のものかを迅速に判断する必要がある。

1852年ロンドンで流行したコレラの発生源を突き止めた John Snow は、患者の居住地を地図上にプロットしてその中心にある水道ポンプに原因があることを示した⁹⁾。このエピソードは疫学調査による地理的な解析が健康危機に有効であることを示した最初の例として有名であるが、多数の患者の居住地を一つずつ地図にプロットする作業には、地図が用意されているとしても長い時間がかかる。しかし SDMS を利用すれば、図2の例では600人以上の住所を地図にするのにほんの数秒しかかからない。

今日の疫学では、地図化した後に各種の数値解析を行い、より詳しい推定を行う必要がある。しかし、まず地図にしてみるというこれまで時間のかかった作業が何の事前準備もなく一瞬で終わるため、手軽におおよその見当をつけることができ、無駄な作業や解析を行って時間を浪費することを防ぐことができる。また、SDMS には変換した点データを GIS にエクスポートする機能もあるので、より詳しい解析は GIS で継続して行うこともできる。

2つめの例は、ウェブ上に公開されている伊丹市交通事故情報の例である。交通事故情報は、1週間ごとにまとめて表形式として、ウェブページに公開されており、発生日時や事故内容とともに発生場所の住所が記載されている。この場合、このウェブページをいったんデスクトップ上に保存して、SDMS にドラッグ&ドロップすればよい。ウェブページを保存するのが難しければ、上部にあるメニューバーから「ファイル」を選び、そのプルダウンメニューの中から「URL 指定」を選択し、ウェブブラウザのようにウェブページの URL を直接入力して、新しいドキュメントとして登録することもできる。これは場所情報を簡単にネットワーク上で共有する方法として利用することもできるが、専用の地図サーバなどは一切不要で、既存のウェブサーバにファイルをアップロードするだけでよい。

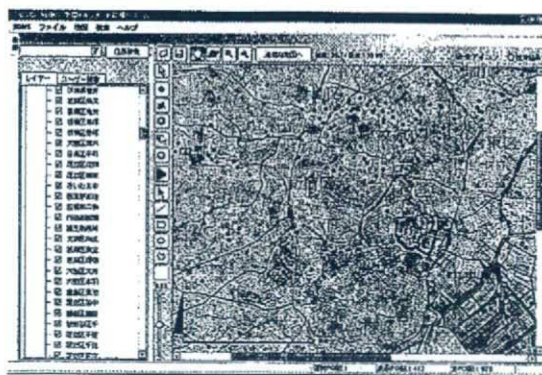


図2. 空間ドキュメント管理システムの東京大気汚染公害裁判原告の住所リストから作成した空間分布図。東京大気汚染公害裁判に関する詳しい説明は以下を参照していただきたい。http://taiki-tokyo.web.infoseek.co.jp/ 本裁判は、2007年8月8日に和解が成立している。

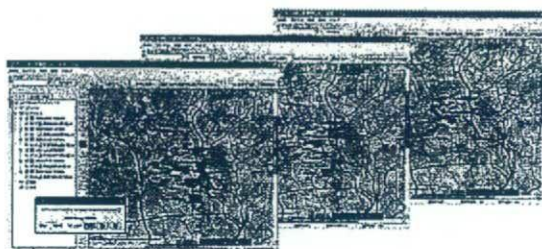


図3. SDMS の時計ブックマーク管理の例。伊丹市交通事故情報のウェブページに掲載してあった一週間ごとの交通事故発生一覧 (6時系列) から空間分布図を作成し、アニメーションで重量表示させた例。(データソース) http://www.itami.fm/seikatu/jiko/index.html

1つのドキュメントは、1つのフォルダとして、左側の空間ブックマークに登録される。図3の例では、それぞれのブックマークを、時計ブックマークにして時系列情報として保存した結果をアニメーションとして再生する様子をイメージ化したものである。

3. SDMS の技術

3.1. 一般ドキュメントから POI への処理手続き

ドキュメント中心の SDMS では、GIS のような数値による場所情報データを持っていない。そのため、一般ドキュメントから POI (Point of Interest) つまり地図上の点地理オブジェクトの自動生成が中心的な情報処理となる。この処理は、主に2つの部分処理、(a) ジオパージング (geoparsing) と (b) ジオコーディング (geocoding) から構成されている。以下、それぞれについて説明する。

3.1.1 ジオパージング

ジオパージングは、一般ドキュメントから住所や地名の抽出を行う処理である。たとえば電子メールの文章から会議場の住所の部分だけを見つけ出して、切り出す処理と考えると分かりやすい。技術的には、自然言語で書かれた文章から場所に関する名詞の部分適切に抽出する処理であり、仮名漢字変換などに用いられている自然言語処理技術の一つである。しかし住所や地名には「田中」や「中村」