

地理疫学情報解析システムの開発と現状

Development of geographical and epidemiological information analysis system

三上 春夫* 高山 喜美子 田中 留美

1. 概要

千葉県がんセンター疫学研究部では 2000 年より継続して地域がん登録データの活用を目的とした地理疫学情報解析システムの開発を行ってきた。システム名称は C3-GEIAS (Chiba Cancer Center-Geographical and Epidemiological Information Analysis System) で、市販の Windows 用地理情報解析ソフトウェア MapInfo(R)をベースとした解析システムが PC 上で動作するように設計した。今回 Ver.3 として(1)住所データを含むファイルから地図上への一括自動マッピング(ジオコード)機能、(2)解析のための関心領域(バッファ)として不定形バッファおよびメッシュバッファの設定、(3)円・扇形・折れ線バッファについて基点からの距離計測機能、(4)人口密度補正機能、(5)広域地図印刷などの機能追加および改善を行った。これらの機能紹介と応用例を提示する。

2. 地理疫学情報解析システム開発の目的と経緯

地域がん登録は本来がん罹患の地域特性を明らかにするものであるが、これまで罹患率や生存率といった指標は市町村などの行政区分を単位としてしか集計不能であり解析には自ずと限界があった。また近年がん罹患と各種環境要因の関連が明らかになるにつれて、定点観測データや人工衛星観測データなど行政区画とは異なったデータ区分を利用する環

境モニタリング的な手法と連携した解析の必要性も出てきた。そこで PC 上での利用が急速に進歩してきた地理情報システム(GIS; Geographical Information System)を採り入れてその有用性を検証するために地理疫学情報の解析ソフトウェアを開発することとした。

3. 開発環境

開発には Microsoft 社のオペレーティングシステム Windows 98 および NT 上で動作する MapInfo 社製地理情報システム MapInfo® Ver5.0 と MapInfo 上の開発環境 MapBasic を使用した。PC は画像データ処理上最低でも PentiumIII 300MHz 以上、RAM128MB 以上を要し、HDD 容量はインストールする地図データの種類と容量にもよるが当疫学研究部で使用中の地図(全国行政界図、関東地方道路図、1市町村住宅地図等)を組み込んだ状態で 900MB 超の容量を占有している。ソフトウェアの改良により動作速度に改善の余地はあるが、大量の画像処理を要する地理情報システムには相当のマシンパワーを消費することは事実である。また地図印刷のためにカラープリンタを用意することが望ましい。

必須ソフトウェアとしては、MapInfo Ver5.0 以上、住所マスタ、行政界地図、住所ジオコードソフトウェア、等高線ライター for MapInfo(R)、使用データとしては国勢調査人口データ、メッシュ地図データ、道路地図データ、町丁目地図データ、住宅地図データ、

*千葉県がんセンター疫学研究部

〒260-8717 千葉市中央区仁戸名町 666-2

ポイントデータがあり、これらは三井造船システム技研と代理店より MapInfo 専用データとして提供されている。

4. 機能説明

開発したソフトウェアは MapInfo の一つのメニューとして動作する。以下操作手順に沿って機能の概要を説明する。

(1) ファイル読込

解析対象のデータファイルは CSV ファイル(コンマ区切りのテキスト)で用意する。ID、住所、診断コード、性別、年齢区分、発症年代などのデータ構成は INI ファイルに記述する。ファイルが読み込まれこれを保存した後「住所ジオコード」を実行する。なお補助的に追加データはマウスで地図をクリックしながら個別入力できる。

(2) 住所ジオコード

住所ジオコードは患者住所を地図上のポイントに落とす(マッピング)機能で、住宅地図と一致すれば号レベル(「～番地～号」すなわち住宅)にポイントを打つことが可能である。住宅地図に掲載のない住所では所定の範囲で番地レベル(すなわち「～丁目～番地」といった区画)にポイントを打つ。それでも一致しない住所(町名変更や慣用的呼称)は解析失敗状況がファイル出力される。通常都市部では 80%以上の成功率が得られる。なお補助的に住所を漢字入力しての個別入力ができる。

(3) 地図表示

マッピングしたデータを、例えば性別や診断区分(ICDコードなど)や年齢階級など INI ファイルに指定した項目とマーカ(図形)により選択表示する。追加地図表示の際には前の表示を残すか消去するか選択できる。この機能により各がんの分布特性などを地図上で確認できる。

(4) バッファ生成

関心領域(バッファ)を生成してデータ(患

者)を抽出する。バッファには円(基点と半径を指定)、扇形(基点と角度と半径を指定)、折れ線バッファ(基点になる折れ線と線からの距離を指定)、不定形バッファ(地図上の観測点と観測データを指定)、メッシュバッファ(JISメッシュ番号と観測データを指定)が用意されており、各バッファ間の和・積・補集合などの演算の結果を新たなバッファに指定できる。

折れ線バッファは道路や河川や送電線などを基準にデータを抽出することを想定している。地図データにより「～市の12時間交通量5000台以上の幹線道路」や「送電設備」などのラインを検索抽出してバッファにすることができる。

不定形バッファは大気汚染や気温など観測点の住所と観測データによりスプライン補間を用いてコンターマップ(等高線図)を描く。また土地の高度や水深など既に作成済みの地図を取り込むことができる。等高線の任意の区間を一つのバッファとしてデータ(患者)の抽出が可能である。

メッシュバッファは人工衛星で観測した植生データや人口密度といったメッシュ単位でデータが与えられているもので、任意の区間に含まれるメッシュを一つのバッファとして扱うことができる。メッシュは1km方形のJIS3次メッシュと500m方形のMapInfo固有4次メッシュが使用できる。

(5) 距離計測

上記のバッファで抽出したデータ(患者情報)を出力する際、円・扇形・折れ線バッファでは基点(基線)からの距離を計測し出力する。これにより「道路から50m圏内」などのデータを集計できる。

(6) 人口密度補正

上記のバッファで抽出したデータ(患者情報)を出力する際、データ(患者)の属するメッシュ区画の人口を付加して出力する。人口密度の補正はExcel上などで行う。

(7) 層別・種別集計

性年齢別や診断コードといった INI ファイルに指定した項目間のクロス集計が可能である。

(8) 印刷

MapInfo での基本的な印刷範囲は表示画面単位に限られるので(スケールは任意) 現在画面を中央とし周囲 8 面を加えた 9 面連続印刷機能を追加した。これにより広範囲の一括

地図印刷が可能である。

5. 今後の課題

様々な主題に本ソフトウェアを適用する際、解析過程における患者データの地図表示は個人情報可視化であり、特別な注意を払う必要がある。そこで個人情報を特定するデータの不可視化機能を付け加えるなど、さらなる改良を加えていきたい。

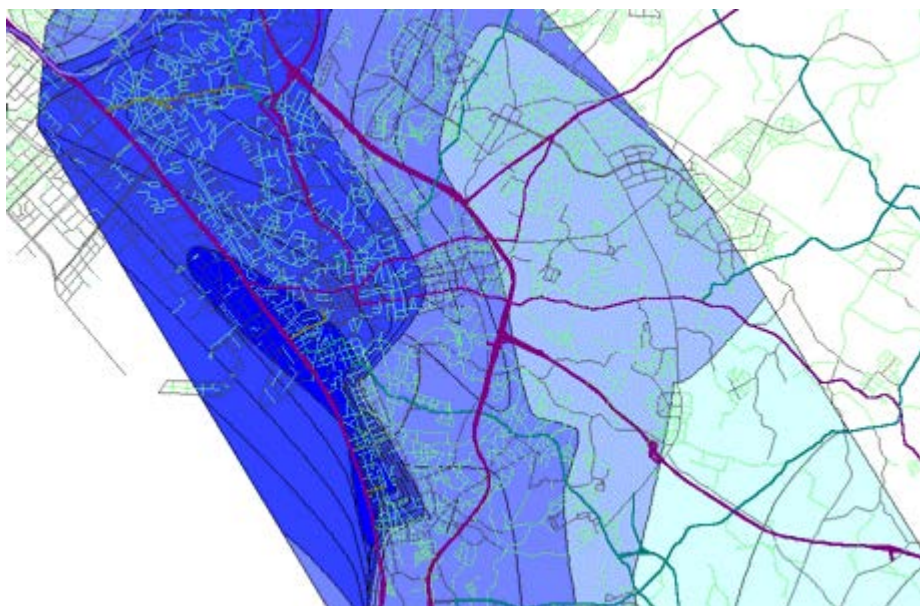


図. 道路地図と大気汚染データ (SO₂) の補間表示