

(財)日本建設情報総合センター研究助成事業

紙面地盤情報電子化技術を利用する地盤情報 DB 構築システムの開発

報告書

平成20年8月

第 1 章 序論

本研究は、H18 年度に実施した紙面地盤情報電子化技術を利用する地盤情報 DB 構築手法を実用化した、システム開発を行うことを主たる目的とする。

そのために、提案手法において課題として残った手書き文字や N 値の判別などに関して地盤工学的知見を利用した認識手法へと高度化するとともに、国土交通省国土計画局国土情報整備室提供の街区レベル位置参照情報を利活用する地盤調査地点の位置情報取得方法を検討する。以上により、位置情報を持った電子化地盤情報となる。

次に、上述した手法により電子化した地盤情報、および、既存の電子地盤情報を用いた、地盤情報 DB における異常値を統計的手法により検出する手法を提案する。ここで、異常値とは、周辺地盤の堆積構造との空間的な相関が極めて低い状況や土質・深度・N 値の組みあわせにおける不整合性のことをいう。すなわち、周りの状況と比較して疑わしいと判断される地盤情報を検出し、該当情報を再検討し、修正あるいは除外することによって、地盤情報 DB のクオリティを向上できる手法を提案する。

以上の手法について、既存の紙面地盤情報について具体的に検討するとともに、実務者が利用しやすい環境を持った紙面地盤情報電子化技術を利用する地盤情報 DB 構築システムの開発を行う。

第2章 地盤情報電子化技術を利用する地盤情報 DB 構築システムの開発趣旨と意義

本研究は、H18 年度に実施した紙面地盤情報電子化技術を利用する地盤情報データベース（以下、地盤情報 DB と略称する）構築手法を適用した地盤情報 DB 構築システム開発を行うことを目的とする。そのために、①電子化技術における認識精度の高度化、②地盤調査の位置情報入力支援、③電子化地盤情報の相互比較による異常値検出と BD の信頼度向上、かつ、④ユーザーフレンドリーな操作環境について詳細な検討を実施し、これらを総合化することによって目的を達成する。

柱状図や各種土質試験をデジタル化した地盤情報データベース（以下、地盤情報 DB と略称する）が構築され、建設時における地盤調査計画、構造物設計などの業務支援や地域の防災および環境保全のための貴重な情報源と位置づけられている。わが国における地盤情報 DB は、とりわけ主要都市部において構築・公開が進んでおり、これは近年の情報電子化による DB への移行の容易さと地盤調査の機会が多さなどに依存している。これに対し、地方都市においては、建設頻度の低さから、調査地点が疎な地盤情報 DB と成らざるを得ない。しかし、過去の地盤調査結果や地盤柱状図集といった紙面情報が存在することが多く、これらを効率的かつ十分な信頼度を持ってデジタル化する本研究は、地盤情

報収集が困難な地域では地盤情報 DB 構築の大きなドライビングフォースとなりうる。加えて、地域の建設業務支援や、防災・環境保全のための貴重な情報整備を支援する技術開発であり、社会的経済的な点と DB の整備状況から見た地域格差の是正の点で意義がある。

第3章 紙面地盤情報を効率的にデジタル化する 技術の精度向上

3.1 はじめに

土質柱状図には、土質区分と深度、標準貫入試験結果など多様な記載情報が記載されており、この土質柱状図画像から適切に各種情報を読み取ることが必要である。土質柱状図の様式は、昭和62年の「地質調査成果資料整理要領(案)」¹⁾以降、ほぼ統一されているが、それ以前は、調査業者独自の様式が使用されている。すなわち、土質柱状図情報の個々の読み取り位置を一般化することは不可能に近いと考えられる。しかし、一冊の報告書を手にとってみると、これはある調査業者によってとりまとめられたものであり、報告書内に収められている土質柱状図の様式は、統一化されている。また、国土交通省地方整備局、都道府県、市町村などの管理単位で考えると、地盤調査を実施している業者には限りがあると予想されることから、ここで開発する技術を用いる際には、土質柱状図の様式を数種

類に限定しうるものと思われる。また、柱状図のヘッダ部に相当する調査名、事業工事名などは、同一の報告書内では同じである。さらに、深度方向の標尺、標高、深度、柱状図、N 値など各種情報は、同一の報告書に収められていたとしても、調査位置によっては異なるのが通常であるが、ここでの記載箇所は、横一線に並べて記載されるのが一般である。このような柱状図特有の性質を利用した読み取り位置決定が有効であると考えられる。

H18 年度に開発した電子化アプリケーションソフトウェアでは、手書きの文字認識や罫線の重なる文字において認識度が低下しており、実用化には不十分であった。それを解決するための文字認識の精度向上が大きな課題であった。

3.2 土質柱状図の特徴に基づいた認識効率化

本研究における電子化技術の第一ステップとしては、まず、紙面情報である土質柱状図をスキャナにより読み取ることから始まる。読み取った画像データから既存資料調査による土質柱状図情報のカテゴリー化として、書式によるカテゴリー化、形式によるカテゴリー化、情報量によるカテゴリー化を行った。

書式によるカテゴリー化では、土質柱状図の記載事項として、紙面の上部に位置する調査位置、調査件名、発注機関などの標題情報と、地表面から深さ方向に関する土質岩種区分、色調、観察記事、標準貫入試験などのいわゆるコア情報から構成される。

次に形式によるカテゴリー化では、認識すべき情報とその種別、および、認識しなくてもよい項目名や罫線などを区別分類することによって、認識率の向上が図られると思われる。例えば、標高値には数字と小数点、符号の数値しか現れない、土質区分には漢字などの文字で記載されているなどである。

そして、情報量によるカテゴリー化では、読み込んだ情報で、相互に整合性を持つべき情報（例えば、標高値、深度、層厚の数値の変化量、土質名称には限られた文字、“年度”という文字は無いが“粘土”はある、など）に着目し、認識結果に対して整合性が要求される項目である。

H18 年度では上記のことに基づいて認識ソフトウェアを開発した。本研究では認識精度向上のために、土質名称のための認識辞書の組み込み、および、罫線や斜線、印刷のシミなどの除去についての工夫を行い、認識精度の向上を行った。

3.3 認識精度向上結果

年度の異なる複数の柱状図印刷物を用意し、H18 年度開発したソフトウェアによる認識結果と本年度改良したソフトウェアによる認識結果の比較を行った。表-3.1 は H18 年度開発したソフトウェアによる認識結果を示している。表中の「無」は原稿には存在している文字が認識されず読み飛ばされたものである。また、「他」は無いものが文字として出てきたもの（例えば罫線とかよごれなど）である。最近の紙面データはワープロ印字や印刷が鮮明であるため認識度は高いことが分かる。一方、年度が遡るにしたがって、認識精度が落ちていくことが分かる。

そこで、以下の2点について改良を行った。

(1) 認識前の画像処理

画像の濃度グラフを表示し、ユーザに読取り範囲を指定してもらうことによって、不明瞭な罫線をはっきりとした線にひきなおす。小さなゴミ、よごれを消す。改良前は「文字」として誤認識されていた、斜線による塗りつぶしを認識対象からはずすといった改良。

(2) 単語辞書の使用

土質、色調欄の認識に柱状図に出てくる単語に特化した辞書を使用するよう改良。このような改良を行うことによって、認識精度の向上を確かめたのが、表-3.2 である。全てのケースで認識精度が向上しており、上記の改良が効果的だったことが分かる。とりわけ、古い柱状図情報で効果的だったことは、この改良が有効だったことを示している。

表-2 改良後のソフトウェアによる認識結果

	標高			深さ			層厚			標準貫入試験												色調			土質			ヘッジ			計			改良後		改良前																											
										深さ			打撃回数/貫入量			10cm			20cm															30cm			正答率	向上率	正答率	向上率																							
	計	正	誤	計	正	誤	計	正	誤	計	正	誤	計	正	誤	計	正	誤	計	正	誤	計	正	誤	計	正	誤	計	正	誤	計	正	誤	計	正	誤	計	正	誤																								
	他	他	他	他	他	他	他	他	他	他	他	他	他	他	他	他	他	他	他	他	他	他	他	他	他	他	他	他	他	他	他	他	他	他	他	他	他	他	他																								
H17 ページ2 画像 0001	11	5	5	1	0	9	5	0	4	0	9	5	4	0	0	226	##	0	4	0	30	25	3	2	0	28	26	1	1	0	28	28	1	1	0	89	81	6	2	1	12	5	7	0	1	14	12	2	0	0	44	44	0	0	500	458	29	15	2	91.20	1.40	89.90	
H15 ページ1 画像 0001	23	19	4	0	0	20	15	0	5	0	20	15	5	0	1	56	56	0	0	2	9	9	0	0	3	10	10	0	0	7	10	10	0	0	3	26	16	10	0	0	16	15	1	0	0	17	17	0	0	1	41	41	0	0	248	223	20	5	17	89.92	10.89	79.03	
H13 ページ4 画像 0001	24	24	0	0	0	16	16	0	0	0	16	12	4	0	0	24	24	0	0	8	7	8	1	0	1	2	2	0	0	3	1	2	0	0	10	8	2	0	3	8	4	4	0	0	11	7	4	0	0	56	50	6	0	0	177	154	23	0	12	87.01	6.21	80.79	
H11 ページ1 画像 0001	26	25	1	0	0	20	18	0	2	3	20	18	2	0	0	60	27	0	33	0	8	8	0	0	9	5	4	0	0	4	2	2	0	0	28	19	9	0	0	13	13	0	0	0	16	14	2	0	0	43	43	0	0	0	248	192	20	35	3	77.42	7.26	70.16	
H09 ページ1 画像 0002	28	22	5	1	2	21	14	2	5	8	20	15	1	4	0	92	72	10	10	9	44	27	15	2	19	23	22	0	1	36	11	11	0	0	44	8	7	1	0	45	22	12	7	3	4	24	13	9	2	3	67	61	4	2	7	293	215	50	28	170	73.38	24.91	48.46
H07-1 ページ 10 画像 0001	17	14	1	2	4	13	7	6	0	3	なし				なし				5	5	0	0	0	なし				なし									6	2	3	1	3	16	13	1	2	0	41	33	5	3	6	57	41	11	5	10	71.93	14.04	57.89				
H05 ページ2 画像 0001	28	2	0	26	0	23	6	4	13	1	20	2	1	17	1	101	58	3	40	5	46	26	8	12	24	21	16	2	3	20	12	9	2	1	28	3	3	0	0	37	10	5	0	15	15	8	7	0	27	35	32	2	1	3	279	135	32	112	158	48.39	4.50	43.88	
H01 ページ1 画像 0001	18	15	1	2	1	14	10	3	1	6	12	9	2	1	0	42	36	4	2	10	13	9	3	1	71	12	8	4	0	22	4	3	1	0	14	4	1	1	2	0	6	3	3	0	0	13	11	2	0	0	34	26	5	3	1	138	105	24	9	124	76.09	33.33	42.75
S57 ページ 01 画像 0001	46	32	10	4	4	33	27	5	0	1	35	30	4	0	1	100	75	12	37	13	26	4	18	0	4	23	13	8	1	2	19	6	12	2	1	22	12	7	0	3	28	7	19	0	2	52	25	21	0	6	62	52	6	0	4	356	248	77	41	31	69.66	16.16	53.50

第4章 地盤調査の位置情報入力支援

4.1 はじめに

地盤情報の電子化において位置情報の精度も同時に要求される。多くの土質柱状図には緯度経度情報欄があるにも拘らず、空白となっている場合が多い。そのため、柱状図認識ソフトウェアによる電子化だけでは対応できないことがほとんどである。

このような場合、柱状図だけでなく、その調査位置を記した調査位置図から緯度経度情報を取得することとなる。既存の手法としては、GIS（地理情報システム）を活用し、例えば2万5千分の1地形図に、調査位置図の画像情報を重ね合わせ、これにより緯度経度座標系と同一にすることにより、調査位置の地点をプロットし、座標値を取得する。GISを用いるとポイントデータとして座標値情報を電子化したまま取得することができ、効率が良い。GISが無い場合は、近年では、インターネット上で表示されるマップを利用する方法があり、この場合は、調査位置の地点図とマップを対比させながら緯度経度情報を記録する。記録された緯度経度情報は手入力によりXMLファイルに書き込むこととなる。その

他、緯度経度情報には、地図上に落とし込む誤差や入力時の記入ミスなど、地盤情報の精度を左右する大きな問題が存在する。特に、入力によるミスは位置が重要な地盤情報システムでは致命的なものであり、地盤情報データベース全体のクオリティを下げると言っても過言ではない。

本章では、既に電子化された地盤情報を用いて、その位置精度の確認手法の確立、修正方法の提案について述べる。

4.2 対象地域と地盤情報DB

対象とした地域は、茨城県である。この地域の地盤に関する情報は、国土交通省 **Kunijiban** に収められているデータと茨城県が管理しているデータベースがあり、本研究ではこれを用いた。図-4.1 に本研究で用いた地盤情報の位置図を示す。図から分かるように、いくつかの地盤調査地点が県外、海上などに存在することが分かる。

4.3 位置確認方法

位置確認方法として、GIS 上で位置情報と市町村ポリゴンから抽出した市町村名 (ADDRESS) とボーリングデータの位置 (住所) が合っているかを確認した。整合が取れているものについて、OK, エリア外, 海上, **ERROR** といった項目に振り分けることにより確認を行った。

エリア外, 海上に位置するデータは合っているデータには「0」、明らかに合っていないデータについては「7」、検討を要する地点には「9」を記した。

ERROR では、合っているデータには「0」、明らかに合っていないデータについては「4」、検討を要する地点には「5」を記した。結果を表-4.1 に記す。また、照合に要した時間を表-4.2 に記す。この位置情報の確認には多大な時間を要することが分かる。このことから、位置情報の取得は十分な注意を払って行うことが必要であることが分かる。

表-4.1 整合結果

項目	番号	データ数
エリア外	7	8
	9	1
海上	7	0
	9	5
ERROR	4	12
	5	6

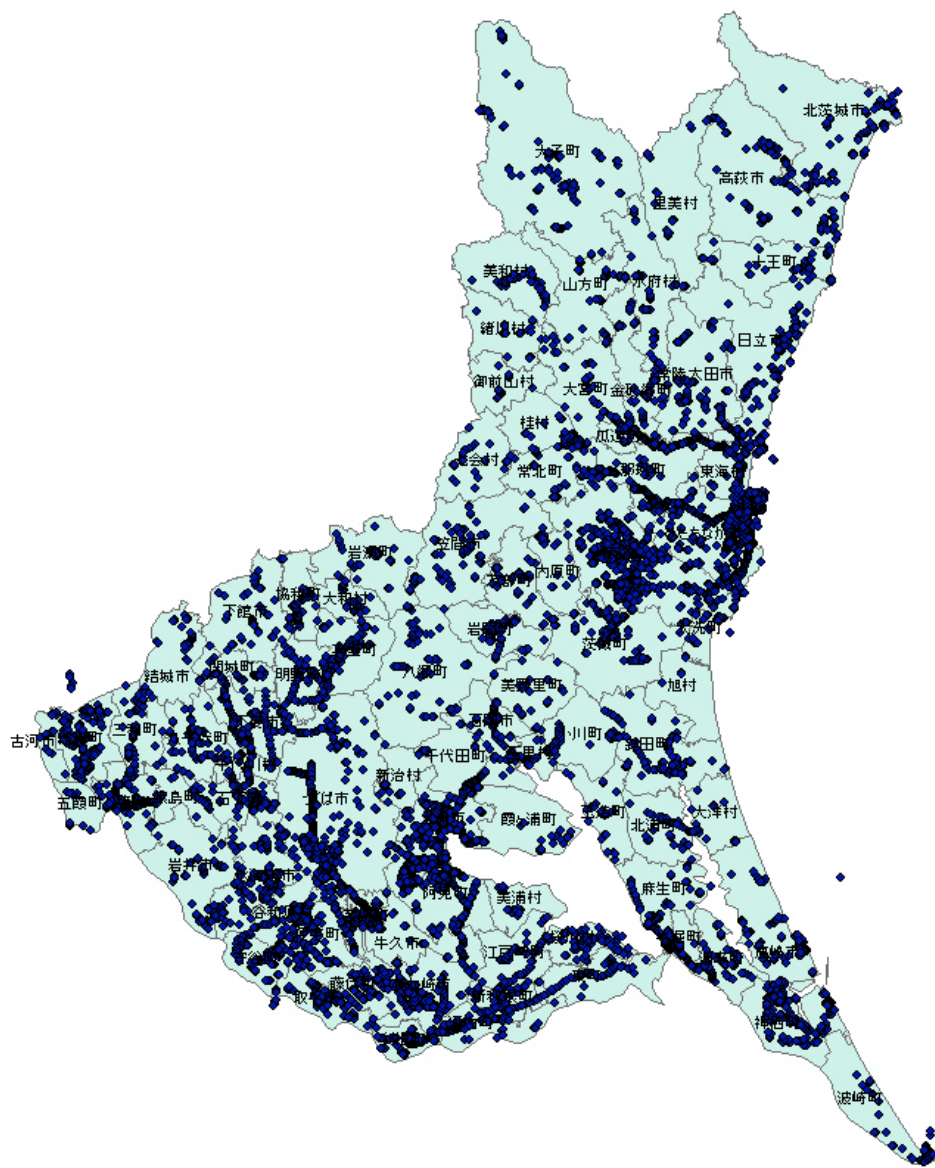


図-4.1 地盤情報位置図

表-4.2 照合に要した時間

項目	総データ数	所要時間 (h)
エリア外	133	6
海上	248	9
ERROR	452	16

4.4 標高値確認方法

上述した位置の確認は平面方向の確認であった。地盤情報は3次元の情報であり、鉛直方向の位置の精度も重要である。50mDEM 標高値と孔口標高値の標高値差を確認し、両者の差が大きい場合にはその値を修正またはチェックすることとした。具体的には、標高値差 40~∞を対象として確認、保留、修正する。CHK3 の欄に、小数点を変えて修正した値に関しては「1」を、孔口標高の値が 0 だった場合に DEM 標高の値に変換したデータに関しては「2」を、確実に値は正しくないと思われるが、修正できなかったデータに関しては「9」を記した。結果を表-4.3.に示す。

表-4.3 標高値の確認結果

CHK3	データ数
1	14
2	32
9	445

「9」の 445 個のデータのうち 314 個は DEM 標高の値が「-1000」と記しており、水域であると考えられる。多くの場合、位置情報が海上にプロットされていた。この作業時間はおよそ 3 時間であった。疑いのある地盤情報については、その周辺の地盤情報との層深度と地層の連続性を照らし合わせ有効な高さに変更するか、あるいは、50mDEM 標高値に置き換えるなどの対処の方法が考えられるが、そもそも、地盤情報を書き換えて良いものかどうか判断できない。個人の利用としては修正も可能と考えるが、一般に公開する場合には、やはり元の調査報告書に立ち戻り標高値を確認することが必要であろう。

第 5 章 紙面地盤情報電子化技術を利用する高密度 地盤情報 DB 構築手法の提案

5.1 構築手法の提案

地盤調査報告書や地盤柱状図集といった既存の紙面地盤情報を、デジタルスキャニング技術を利用し、個々の情報を効率的に電子化するとともに、集積したデジタル地盤情報から個々の情報の信頼度を地盤統計学的手法に基づき判定するとともに、高密度地盤情報データベースを構築する手法を開発する試みを行った先の研究成果を更に発展させるため、読み取り精度の向上、電子化された地盤情報の確認方法について、実データを用いた実用化の検討を行った。

以上の成果を有機的に結合させることによって紙面柱状図を効率的にまた大量に電子化することによる高密度地盤情報データベース構築できる。その手法を図 5.1 に示す。

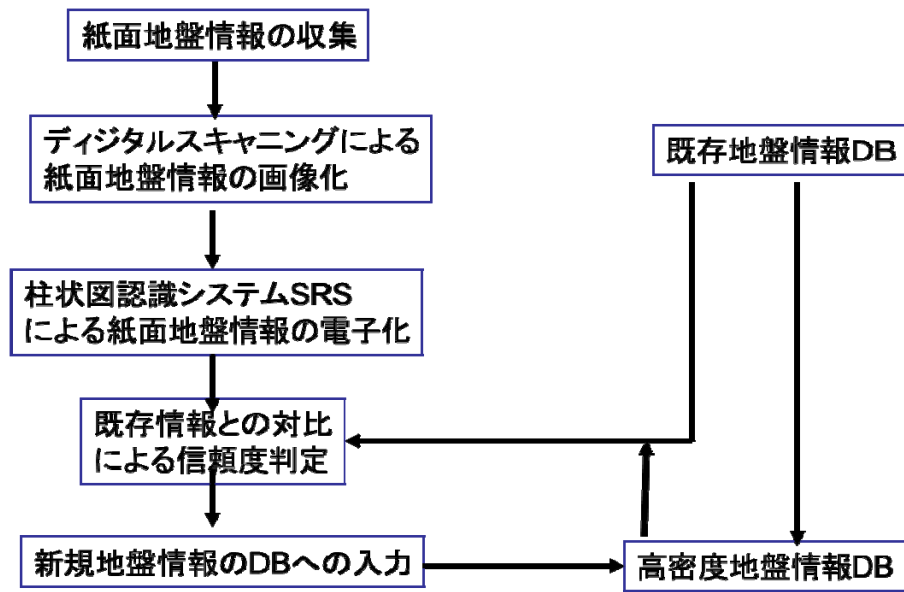


図 5.1 紙面地盤情報電子化技術を利用する高密度地盤情報 DB 構築手法

助成研究者紹介

むらかみ さとし

村上 哲

現職：茨城大学工学部講師（工学博士）

うえだ よしかず

上田 賀一

現職：茨城大学工学部准教授（工学博士）