

論文

入学志願者増加を目指す広報活動のための オープンデータを活用したデータ分析

高橋 章¹・上村 健二¹・佐藤 拓史¹・佐藤 秀一²

¹ 電子制御工学科 (Dept. of Electronic Control Engineering, National Institute of Technology, Nagaoka College)

² 一般教育科－物理 (Liberal Arts-Physics, National Institute of Technology, Nagaoka College)

Data Analysis for Publicity Campaign Aiming at Increasing in Candidates for Admission Utilizing Open Data

Akira TAKAHASHI¹, Kenji KAMIMURA¹, Hiroshi SATOH¹, and Shuichi SATO²

Abstract

This paper describes data analysis for publicity campaign aiming at increasing in candidates for admission of National Institute of Technology, Nagaoka College. Both the efficiency and the accuracy of reception procedure for Open Campus are increased by utilizing open data published on website of Niigata prefecture, and summary tables of junior high schools and municipalities for participants are generated automatically. In addition, databases of junior high schools are constructed by gathering open data, record of entry for Open Campus, record of admission, and so on. Finally, as an example of data mining, the result of promotion campaign to visit junior high schools selected from the databases is shown.

Key Words : *open data, data mining, database, spreadsheet, spreadsheet function*

1. はじめに

情報通信技術 (ICT; Information and Communication Technology) の進展により、インターネット上で教育や商取引などを行うための様々なサービスが開発され、利用が広がっている。膨大な顧客データや購買履歴などを収集してデータベースを構築し、それらを分析して動向の予測等につなげるデータサイエンスや、個々のユーザへのきめ細やかな対応や新たな商品・サービス開発に役立つ付加価値を創造したり意思決定に活用したりするデータマイニングが盛んに行われており¹⁾、これらを総称してビッグデータと呼ぶことも多い。国や自治体が様々なデータを Web 上で公開するオープンデータと呼ばれる行政サ

ービスも推進されている。

高等専門学校の学校経営における至上命題の一つとして、中学校からの入学志願者を安定確保することがあげられる。長岡工業高等専門学校（以下、本校）においても、平成 25 年度に行われた学内組織改革により入学者対策室と広報委員会が一元化され、広報戦略室として志願者を増やすための広報活動を担うこととなった。その代表的な活動として、中学校で開催される高校説明会における学校紹介、本校教員による中学校訪問の企画、オープンキャンパスおよび入試説明会の計画・実施などがある。従来、これらの活動に付随して作成されるデータは、学生課入試係で年度およびイベント毎に記録・管理されていた。より戦略的な広報活動を行うには、個別に

	A	B	C	D	E	F
1	平成28年度新潟県市町村立中学校(県内全体)					
2	No	市町村	学 校 名	郵便番号	所 在 地	電話番号
3	1	上越市	城北中学校	943-0827	上越市	025(523)
4	2	上越市	城東中学校	943-0835	上越市	025(522)
5	3	上越市	城西中学校	943-0847	上越市	025(523)
6	4	上越市	雄志中学校	943-0119	上越市	025(523)
7	5	上越市	八千浦中学校	942-0022	上越市	025(543)
8	6	上越市	直江津中学校	942-0004	上越市	025(543)
9	7	上越市	直江津東中学校	942-0041	上越市	025(543)
10	8	上越市	春日中学校	943-0803	上越市	025(522)
11	9	上越市	潮陵中学校	949-1711	上越市	025(546)
12	10	上越市	安塚中学校	942-0404	上越市	025(592)
13	11	上越市	浦川原中学校	942-0314	上越市	025(599)
14	12	上越市	大島中学校	942-1105	上越市	025(594)
15	13	上越市	牧中学校	943-0648	上越市	025(533)

図-1 市町村立中学校（抜粋・一部加工）³⁾

作成されるデータを適宜集約して分析することにより、経年的な変動や、イベント間の参加者の推移などの動向を、迅速かつ正確に把握することが求められる。しかし、そのようなデータ分析を行うには、作表ソフトウェアやデータベース管理に関する知識とスキルばかりでなく、戦略的な視点が不可欠であり、人的にも時間的にもコストが高いという課題があった。

本報告では、新潟県が Web で公開しているオープンデータ²⁾を活用し、オープンキャンパスの受付作業における労力を軽減し、集計作業の迅速化および精度向上をはかるためのデータ処理について述べる。さらに、県内の公立中学校の生徒数や、本校との関わりをデータベース化して行った分析について述べる。これまでオープンキャンパスへの参加者がなく、出願者のいなかった中学校へ訪問活動を行うことによって、効果的に PR 活動の成果につながったデータマイニングの事例を示す。

2. オープンキャンパス受付・集計作業

オープンキャンパス（OC）は中学生とその保護者を対象とする、本校最大の PR イベントである。平成 25 年度以降、参加者は増加傾向にあり、7 月の一ヶ月間で 400 名を超える中学生からの参加申し込みを受け付け、8 月上旬の数日間体験学習の割り付けを行い、参加者への通知文書の発送を行う。参加申し込みの手段には郵送、FAX、電子メールなどがあり、その大部分は手作業で受付簿に入力しなければならないため、人的コストを低減し、入力ミス等を防ぐ対策が望まれている。また、OC 終了後の入試説明会、秋季中学校訪問などのイベントの計画・実施を円滑に進めるために、中学校別や市町村

別の参加状況を迅速かつ的確に把握することが望ましい。しかし、従来は、国立大学附属の中学校名を入力する場合の大学名の省略形や「附属/付属」の表記の揺れ、市町村立の中学校名を入力する場合に間に「立」を加えるかどうかといった表記の揺れが混入しやすく、集計する際の確認作業に手間がかかり、精度にも問題があった。

そこでオープンデータを活用し、参加者の所属中学校の入力作業の労力を軽減し、集計作業の精度向上を図る。

2. 1 中学校一覧

新潟県が公開するオープンデータのうち、市町村教育委員会・学校一覧（中学校一覧）は、県内の市町村立中学校 230 校を一覧表にした Microsoft Excel 形式のファイルである³⁾。図-1 のように、印刷資料としても利用できる書式設定が行われているものの、2 行目以降は 1 行 1 レコードとして扱えるデータベース形式となっている。

2 行目はデータベースの見出し行で、1 レコードが A 列～F 列の 6 フィールドからなり、それぞれ、No（学番）、市町村、学校名、郵便番号、所在地、電話番号が格納されていることを示す。学番は上越地区、中越地区、下越地区の順に付けられているが、同一市町村内の学校名の順に規則性はみられない。

なお、ファイルは毎年、年度当初に更新されるが、学校の統廃合が行われた年には学番が変化するため、注意が必要である。また、国立大学の附属中学校は掲載されておらず、県立および市立の中等教育学校（中高一貫校）は平成 27 年度は記載されていたが、平成 28 年度では省かれているため、毎年確認と手動の追加作業が必要である。

2. 2 オープンキャンパスの受付簿作成

OC の受付簿として、参加者の情報を入力するシート **ES** と、中学校一覧を含むシート **SL** の 2 シートからなるファイルを作成する。シート **SL** には中学校一覧の内容をコピーする。受付作業者が、OC 参加申込書から参加者の情報をシート **ES** に入力する際に、在籍中学校の情報をシート **SL** のデータベースから検索機能によって自動入力することで、入力作業の軽減と表記の揺れや誤変換の混入を予防する。

シート **ES** における学番、市町村、中学校を入力する 3 フィールドを順に **G** 列、**H** 列、**I** 列とする。学番フィールドのセル **g2** に、シート **SL** のデータベースに登録された在籍中学校の学番を入力するだけで、その中学校の市町村名を、セル **h2** に自動入力させるためには、次の関数式を入力すればよい⁵⁾：

=vlookup(\$g2, SL!\$A:\$C, 2, false)

ここで、**vlookup** 関数は 4 つの引数を持つデータベース関数である。第 1 引数は検索値を、列を絶対参照とする **\$g2** と指定している。第 2 引数は検索範囲を、**SL!\$A:\$C** として、シート **SL** の学番、市町村、学校名が含まれる **A** 列～**C** 列を絶対参照により指定している。**vlookup** 関数では検索範囲の左端の列 (**A** 列) から検索値を探す。第 3 引数は列番号として、検索値が見つかった場合にそのレコード (行) の何列目のデータを取得するかを指定するが、今回は 2 列目 (**B** 列の市町村) としている。第 4 引数は検索の型を指定するが、**false** の場合は検索値と厳密に一致するレコードのみを検索し、見つからなければエラー値を返す。**true** または第 4 引数を省略した場合には、検索値の近似値 (検索値未満の最大値) を許容するが、これを指定する際には検索範囲の左端の列は昇順に並べておく必要がある。

ところで、前述の関数式ではセル **g2** が未入力の場合に、**h2** にはエラー値 **#N/A** が入力されてしまう。これを避けるためには **if** 関数でラップすればよい：

=if(\$g2="", "", vlookup(引数は省略))

ここで、**if** 関数は第 1 引数の条件式が真であるときは第 2 引数、偽であるときは第 3 引数を返す。よって、前述の関数式ではセル **g2** が未入力 (空文字) の場合は、**h2** も空文字としておき、そうでなければ **vlookup** 関数による検索結果を自動入力する。

	A	B	C
1	地区	市町村	参加人数
2	下越・佐渡		112
3		佐渡市	0
4		新潟市	67
5		新発田市	8
6		村上市	2
7		燕市	18
8		五泉市	4
9		阿賀野市	6
10		胎内市	3
11		聖籠町	3
12		関川村	0
13		粟島浦村	0
14		弥彦村	1
15		阿賀町	0
16	中越		264
17		加茂市	4
18		三条市	28
19		長岡市	156
20		柏崎市	21
21		小千谷市	15
22		十日町市	5
23		見附市	23
24		魚沼市	6
25		南魚沼市	3
26		田上町	1
27		出雲崎町	0
28		湯沢町	0
29		津南町	1
30		刈羽村	1
31	上越		27
32		上越市	20
33		糸魚川市	2
34		妙高市	5
35	合計		403

図-2 市町村別オープンキャンパス参加者集計表

セル **i2** には、セル **h2** の関数式をコピーした上で、**vlookup** 関数の第 3 引数を **3** として、学校名を取得するように変更する。そして、セル範囲 **h2:i2** の関数式を、下方のセルにオートコピーすることで、受付簿の準備が整う。

なお、県外からの参加申込みがあった場合は、シート **SL** に適宜中学校情報を追加すればよい。

2. 3 オープンキャンパスの集計自動化

中学校別の参加者数を集計するために、シート **SL** に参加者フィールド (**E** 列) を追加する。そして、セル **e3** に次の関数式を入力する：

=countif(ES!\$G:\$G, a3)

すなわち、シート **ES** の **G** 列中で、セル **a3** と同じ学番の個数を数え上げる。

次に市町村別の集計シート **ML** (図-2) を用意し、地区、市町村、参加人数を入力する 3 フィールドを順に **A** 列、**B** 列、**C** 列とする。市町村ごとの参加者を求めるため、当初はセル **c3** に次の関数式を入力すればよいと考えた：

=sumif(SL!\$B:\$B, b3, SL!\$E:\$E)

すなわち、シートSLのB列の値が、セルb3と一致するレコードについて、E列の値の合計を求める。しかし、期待通りに動作しないためシートSLのデータベースを精査したところ、市町村フィールド（B列）の文字列の最後に全角空白が1文字だけ加わっているレコードが存在し、セルb3と不一致として処理されていることがわかった。一つ一つのセルの値を調べて手作業で修正するのは煩雑であるため、シートSLに検索・照合用の市町村名フィールド（F列）を追加し、セルf3に次の関数式を入力する：

```
=if(right(b3,1)="",left(b3,len(b3)-1),b3)
```

すなわち、セルb3の市町村名の右端から1文字目をright(b3,1)によって取り出し、それが全角空白かどうかをif関数により場合分けする。真（全角空白）ならば市町村名の右端1文字分を取り除いた文字列をleft(b3,len(b3)-1)によって切り出す。偽（全角空白ではない）ならば市町村名をそのまま取り出す。4行目以降へオートフィルによって関数式をコピーすることで、全てのレコードに検索・照合用の市町村名が用意できる。

そして、シートMLのセルc3には次の関数式を入力する：

```
=sumif(SL!$F:$F, b3, SL!$E:$E)
```

4行目以降へオートフィルによって関数式をコピーすることで、全てのレコードに市町村毎の参加者数を求めることができる。

シートMLの地区別集計（2, 16, 31行目）と合計（35行目）は、subtotal関数を用いると保守管理が容易となる。セルc2には次の関数式を入力する：

```
=subtotal(9, c3:c15)
```

ここで、subtotal関数の第1引数である9は、集計方法として合計値を求める指定である。第2引数は集計範囲を指定する。セルc16, c31も同様に関数式を入力する。そして、合計を求めるセルc35には次の関数式を入力する：

```
=subtotal(9, c2:c34)
```

すると、subtotal関数によって小計が求められているセル（c2, c16, c31）を自動的に省いた合計値が計算される。

以上の作業により、集計はすべてオンタイムで実行され、受付処理の任意の時点で最新の状況を正確に把握することができる。

都道府県別の参加者数について、約95%が新潟県内である現状では、手動で計数を行っても大きな労力ではない。自動的に求めるならばシートSLで、各中学校の都道府県を入力するフィールドを用意して、市町村別の集計と同様の処理を行えばよい。

3. 中学校訪問のための分析資料と活用

本校では毎年秋季に県内の中学校を訪問するPR活動を実施している。しかし、県土が広く中学校数が多いため、全ての中学校を訪問するのは困難であり、毎年百数十校を選定して訪問活動を行っている。最大限の効果を狙って訪問先の絞込みを行うために、各中学校の生徒数を調べ、過去数年間の高校説明会への参加（派遣）実績、オープンキャンパス参加者数、志願者数、中学校訪問の実績を加えた分析用データベースを作成する。

3. 1 中学校別生徒数

中学校学校別学級数・生徒数は、県教育委員会がまとめる学校基本調査関連データの一部として公開されているMicrosoft Excel形式のファイルである⁴⁾。

図-3のように印刷資料用に加工された一覧表で、市町村立の中学校ばかりでなく、国立大学の附属中学校、県立および市立の中高一貫校の各校について、公開前年度の学級数や生徒数が詳細にまとめられている。しかし、集計（小計）行や空白行が挿入されているため、1行を1レコードとするデータベースとして直接扱うことはできない。特に「第一」、「東」などの複数の市町村に存在する中学校名は、単一行だけでは識別できない。市町村の掲載順序も新潟市、長岡市、三条市、柏崎市、新発田市、...となっており規則性が読み取れず、同一市町村内の中学校の掲載順も中学校一覧³⁾とは必ずしも一致しない。

3. 2 データベース作成

表計算ソフトウェアで新規ファイルを作成し、中学校一覧³⁾をシートSLとしてコピーする。次に、このファイルに中学校別生徒数⁴⁾をシート名SNとしてコピーする。シートSLに新たに生徒数フィールド（G列）を設け、シートSNより検索することとする。ここで、両シートの間で中学校名の検索・照合を行うための一意なキーとなるフィールドが存

入学志願者増加を目指す広報活動のためのオープンデータを活用したデータ分析

	A	B	C	D	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	
1	中学校 学校別学級数・生徒数(平成27年5月1日現在)																						
2																							
3																							
4	学校名 (内は分校名)	本校	分校	へき地指定	総数	学 級 数										生 徒 数							
1年						2年	3年	2個学年	3個学年	知的障害	肢体不自由	病弱・身体虚弱	弱視	聴覚	言語障害	盲・ろう・両目障害・併存障害	総数	1年	2年	3年			
5	総 計	237	1	(33)	2,382	635	666	673	4	-	200	6	4	3	1	-	190	60,358	19,212	20,514	20,632		
6																							
7	国立計	3		(-)	27	9	9	9	-	-	-	-	-	-	-	-	-	1,078	363	362	353		
8	新潟大学附属新潟				9	3	3	3	-	-	-	-	-	-	-	-	-	355	120	119	116		
9	新潟大学附属長岡				9	3	3	3	-	-	-	-	-	-	-	-	-	359	120	123	116		
10	上越教育大学附属				9	3	3	3	-	-	-	-	-	-	-	-	-	364	123	120	121		
11																							
12	公立計	230	1	(33)	2,333	619	649	657	4	-	200	6	4	3	1	-	190	58,582	18,621	19,893	20,068		
13																							
14	県立計	1		(-)	3	1	1	1	-	-	-	-	-	-	-	-	-	64	14	22	28		
15	阿賀野明				3	1	1	1	-	-	-	-	-	-	-	-	-	64	14	22	28		
16																							
17	市立計	218	1	(31)	2,244	598	625	634	3	-	190	6	4	3	1	-	180	56,607	18,014	19,176	19,417		
18																							
19	新潟市	56	1	(-)	726	206	211	211	1	-	45	1	2	-	1	-	48	19,934	6,450	6,728	6,756		
20	関屋				17	5	5	5	-	-	1	-	-	-	-	-	1	494	170	150	174		
21	鳥屋野				25	7	8	8	-	-	1	-	-	-	-	-	1	802	248	272	282		
22	白新				13	2	3	3	-	-	1	-	1	-	1	-	2	233	70	77	86		
23	寄居				12	3	3	3	-	-	1	1	-	-	-	-	1	263	83	96	84		
24	宮浦				19	6	6	6	-	-	-	-	-	-	-	-	1	528	157	184	187		
25	東新潟				18	5	6	5	-	-	1	-	-	-	-	-	1	535	154	204	177		
26	山の下				16	4	5	5	-	-	1	-	-	-	-	-	1	462	137	164	161		
27																							

図-3 中学校学校別学級数・生徒数(抜粋) ⁴⁾

在しないため、それぞれのシートにおいて検索・照合用フィールドを新たに作成する必要がある。検索・照合キーは、市町村名と中学校名を結合した文字列として作成する。

■シート SL における検索・照合キー作成 検索・照合キーを新たなフィールド(H列)に作成する。データベースの先頭行(3行目)のセルh3に次の関数式を入力する：

```
=if(right(b3,1)="",
left(b3,len(b3)-1)&c3, b3&c3)
```

すなわち、市町村名(B列)の右端から1文字目をright(b3,1)によって取り出し、それが全角空白かどうかをif関数により場合分けする。真(全角空白)ならば市町村名(B列)の右端1文字分を取り除いた文字列をleft(b3,len(b3)-1)によって切り出して、中学校名(C列)を結合する。偽(全角空白ではない)ならば市町村名(B列)と中学校名(C列)を単純に結合する。最後に4行目以降へオートフィルによって関数式をコピーすることで、全てのレコードに検索・照合キーが作成される(2.3参照)。

■シート SN への市町村フィールド作成 各行に市町村を識別するフィールドが存在しないため、これを新たなフィールド(Y列)に作成する。図-3のA列を調べると、市町村立中学校は20行目～310行目に入力されており、空行と市町村名の後に、その市町村の中学校名が列挙されている。しかし、3文字以下の市町村名には、見た目の体裁を整えるために文字間に空白文字が挿入されている。このため、関数を組み合わせて、この市町村名を自動的に以降の行へ入力することは非効率的と考えられる。そこで、

やや煩雑ではあるが、市町村ごとに先頭行のセルへ手入力または他のシートからコピーを行い、次行以降へオートフィルによってコピーする。

■シート SN における検索・照合キー作成 検索・照合キーを新たなフィールド(X列)に作成する。前述の作業で市町村フィールド(Y列)が作成済みであるので、キーを作成したい先頭行(21行目)のセルx21に次の関数式を入力する：

```
=if(y21="", "", y21&a21&"中学校")
```

すなわち、Y列に市町村名が記入されていないかどうかをif関数により場合分けする。真(無記入)ならばキーを生成しない。偽(市町村名あり)ならばY列とA列を結合し、「中学校」を加えたキーを生成する。最後に22行目～310行目へオートフィルによって関数式をコピーすることで、中学校名が書かれた全ての行に検索・照合キーが作成される。

■シート SL への生徒数フィールド作成 ここまでの作業で、シート SL とシート SN の各行(レコード)間で一意な対応付けを行うための検索・照合キーが作成されている。これを利用してシート SN の生徒数(総数)を、シート SL に作成した生徒数フィールド(G列)の各セルへコピーしてくればよい。しかし、2.2で用いたvlookup関数は、対象範囲の左端の列をキーとしなければならない、汎用的ではない。vlookup関数の仕様に合わせて、シート SN のオリジナルデータ部分に検索・照合キーのための列を挿入することも考えられるが、次年度以降のデータ更新における作業性が低下する。そこで、データベースの先頭行(2行目)のセルg3に次の関数式を入力する：

=index(SN!\$T:\$X, match(h3, SN!\$X:\$X, 0), 1)

ここで index 関数は第 1 引数で指定されるセル範囲から、第 2 引数で指定される行番号と、第 3 引数で指定される列番号のセルの値を取り出す。すなわち、シート SN で検索・照合キーが入力されているセル範囲 SN!\$X:\$X から、h3 セルと一致する行番号を match 関数で検索する。そして、セル範囲 SN!\$T:\$X の 1 列目となる生徒数（総数）の値を取得する。

3. 3 春季中学校訪問の実施

前節のデータベースに過去数年間の高校説明会への参加（派遣）状況、OC への参加人数、入試における志願者数、中学校訪問の実績を加え、各中学校と本校との関係性を調べた。すると、平成 25～26 年度の過去 2 年間志願者がゼロの県内市町村立中学校が 85 校あり、うち 31 校は全校生徒数が 180 名以上と比較的規模の大きな中学校であることがわかった。そこで、12 市町村から比較的生徒数の多い 20 校を選定し、平成 27 年の 6 月～7 月にかけて、校長と副校長による春季中学校訪問を実施した。

その結果、春季訪問先の 75%となる 15 校から 35 名が OC に参加した。そして、春季訪問先の 60%となる 12 校から 26 名の志願者があった。OC 参加者に対する志願者数の割合を志願率として、全体と比較すると表-1 となる。これまで本校への志願者のいなかった中学校へ、効果的に PR 活動を行うことができたことが推察される。

4. まとめ

長岡高専広報戦略室における入学志願者増加を目指す活動を推進するために、表計算ソフトウェアの関数を利用したデータベース管理機能を活用する取り組みを実施した。新潟県が Web で公開しているオープンデータのうち、公立中学校の一覧表を利用することで、オープンキャンパスの受付簿作成の労力を一部軽減し、オンタイムでの集計を実現できた。

公立中学校の一覧表に、在籍生徒数のオープンデータを組み合わせ、さらに長岡高専広報戦略室における中学校向け PR 活動によって派生した活動実績のデータを加えることで、データベースを作成した。このデータベースをもとに分析を行い、過去 2 年間連続でオープンキャンパスへの参加者がなく、入学志願者もいなかった中学校を 20 校抽出し、春季に訪問活動を行ったところ、その 75%となる 15 校からオ

表-1 OC 参加者と志願者数（平成 27 年度）

区分	OC 参加者	志願者	志願率
全体	416	295	71%
春季訪問先	35	26	74%

ープンキャンパスへの参加があった。さらにその 60%となる 12 校から、入学試験への出願があった。これまで個別に集計し、記録として残すだけであったデータ群から、新たな付加価値を創造するデータマイニングが実現できたといえよう。

データ処理を定型処理化することは、労力の軽減やミスの防止という大きな意義を持つと考えられるが、ややもすると作業内容がブラックボックス化し、保守管理作業が形骸化することにつながりかねない。より戦略的かつ効率的に PR 活動を推進するための指針となる科学的なデータ分析を目指し、PDCA を意識してデータベースやデータ処理方法の改善活動を継続する必要がある。

謝辞：中学校向け広報活動に実際に携わっていただいた渡邊校長をはじめ広報戦略室メンバーおよび学生課・入試係の皆さま、オープンキャンパスや中学校訪問にご協力いただいた教職員の皆さまに感謝します。また、総合情報処理センタースタッフの皆さまにはデータ処理について多くのアドバイスをいただきました。

参考文献・URL

- 1) イアン・エアーズ: その数学が戦略を決める, 文藝春秋, 2010.
- 2) 新潟県: 統計ボックス・オープンデータ,
http://www.pref.niigata.lg.jp/tokei_od.html
- 3) 新潟県: 市町村教育委員会・学校一覧,
<http://www.pref.niigata.lg.jp/gimukyoiku/1196007390231.html>
- 4) 新潟県: 平成 27 年度学校基本調査関連データ, 中学校学校別学級数・生徒数,
<http://www.pref.niigata.lg.jp/kyoikusomu/1356826766000.html>
- 5) 高橋慈子: Excel2013 全機能 Bible, 技術評論社, 2013.

(2016. 10. 3 受付)