

学校サイト更新履歴に関する統計的検討

豊福 晋平
Shimpei Toyofuku

国際大学グローバル・コミュニケーション・センター
Center for Global Communications, International University of Japan

<あらまし> i-learn.jp「日本の学校」では2000年以降学校サイトのサーバ情報やサイト内容を機械的に収集し、これをもとに学校サイト更新率(サイト活性度)を計算・公表している。本研究では、2003年以降のデータを対象として経年変化・月変動を明らかにするとともに、年間更新日数の実績順分布に対する曲線推定を行い、べき乗則に適合する事を確かめた。また、自治体別の分析から、教育委員会からの働きかけが実績中位群に有効に作用することが示唆された。

<キーワード> 学校広報 学校サイト 更新頻度 自治体統計 DDDM

1. はじめに

学校サイトの運用状況を客観的に把握するうえで、更新頻度統計はきわめて有効な情報になりうる。i-learn.jp サイトでは2000年9月以降データベース登録された学校サイトURLに対して、クローラを用いた巡回を行い、持続的に各サイトの更新履歴を取得・蓄積している(豊福2002)。

豊福(2005)は、学校サイトの運用実績の一般的傾向として、きわめて更新実績の高い一部サイトと更新実績の低い大多数のサイトで構成されており、二極化が著しいこと、さらに、更新実績の高い順から並べると、べき乗則の曲線に従うことを明らかにした。

i-learn.jp¹では、通常学校単位の更新状況とランク付けを中心とした情報提供を行っているが、本研究では、特に、更新頻度統計の全体傾向と市区町村自治体単位の集計比較に注目し、それぞれの検討考察を行うものである。

なお本研究では統計処理にPASW(SPSS) 18を用いた。

2. データ取得方法と形式

先に示した通り、本研究で用いるデータは、定期的なクローラ巡回による更新履歴の取得をベースとしている。方法としては、RSS(Rich Site Summary)を配信するサイトからは直接RSSファイルから更新履歴を取得する、RSSに対応しないサイトは、サーバ側

が応答する最終更新日時(Modified)・データ長(File Length)・取得データの要約文字行列(MD5)の差分を求めることで、更新の有無を検知する、各サイト側から通知されるBlog Pingを受信して最新のRSSを取得する、の3通りがある。

現在は の差分解析を必要とするサイトが多く、更新履歴の精度がクローラの巡回頻度(通常は12時間おき)に依存するため、i-learn.jpでは、更新履歴の統計を日付単位に丸めて扱っている。つまり、ある日に1回更新しても20回更新しても、i-learn.jp上では1(その日に更新があったこと)だけがカウントされる。

なお、現状 または のRSS以外で外部のウェブサーバの更新履歴を正確に取得する方法はなく、RSSの搭載自体もウェブサーバの標準仕様になっていない。このため、外部の多様なサーバ仕様によって更新履歴の取得精度には相当のばらつきがあることを前提として見る必要がある。

3. 更新履歴の時系列変化

学校サイト更新履歴のうち2003年4月以降の時系列データを学校種別に示したものが図1である。縦軸は一ヶ月間に更新された回数(日数)の総和である。月ごとの変動はあるものの、ほぼ一定の割合で漸増傾向にある。ちなみに2010年4月での更新回数総和は

小学校 61289 回 (1 校平均 2.78 回), 中学校 32963 回(1 校平均 3.05 回), 高等学校 32045 回 (1 校平均 6.26 回) である。

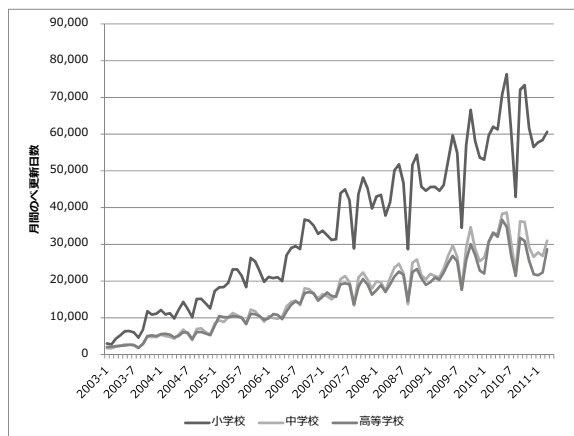


図 1 学校種別月間のべ更新回数の推移

これらのデータを重回帰分析に投入し、ステップワイズ法によりモデル式を得たところ、決定係数 $R^2 = 0.853$, 月間更新回数総和 = $-5214.14(\text{定数}) + 424.33 \times \text{累積月数} + 18068.75 \times \text{小学校ダミー変数}$ となった。

中学・高校のダミー変数はモデルから除外されたので、中学校・高等学校間には更新実績の差がないということになる。

表 1 月間更新回数総和の重回帰分析

	非標準化係数		標準化係数
	B	標準誤差	β
(定数)	-5214.14	766.79	
累積月数	424.33	12.56	0.756
小学校ダミー変数	18068.75	761.63	0.531

4. 更新履歴の月変動

更新履歴の月変動傾向を明らかにするため、2010年4月から2011年3月のデータを抽出し、学校種別に年度平均と標準偏差を求め表2に示した。さらに、この数値を元に年度内の変動を標準偏差で表したグラフが図2である。

表 2 2010 年度月間平均更新回数

	小学校	中学校	高等学校
平均	62693.83	31370.25	27875.42
標準偏差	8829.63	4928.74	5184.97

いずれの学校種でも 5~6 月・9~10 月はプラス傾向、8 月・12~3 月はマイナス傾向が著しい。

特に、小学校について中高と比較すると 3~5 月は値が低めであり、8 月の落ち込みは最も著しいが、一方、10~2 月は中高の値の落ち込み比べるとやや高めである。

これら増減傾向はいずれも繁忙期・長期休業・学校行事との関連が推測される。

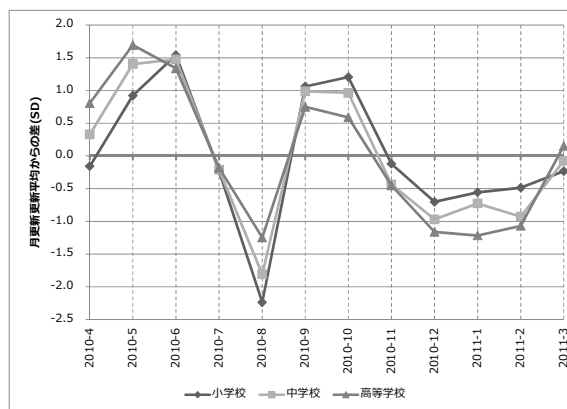


図 2 2010 年度内の月変動推移

5. 年間更新回数のべき乗則分布

図 3は 2004・2006・2008・2010 年の小学校から高校までの全学校の年間更新回数(日数)実績を降順にプロットしたものである。

豊福(2002)および豊福(2005)で述べた通り、この分布はべき乗則におよそ沿ったものを見ることができる。

つまり、きわめて更新頻度の高い学校サイトがごく一部にある一方で、大半の学校サイトの更新頻度は著しく低い、きわめて偏った分布である。

ただし、年を追うごとに年間 200~50 日の領域に分布が遷移しており、中位層にあたる学校の実績が向上しているの、将来的にはべき乗則が崩れ、直線かもしくは逆 S 字を描く分布に変わることが予想される。

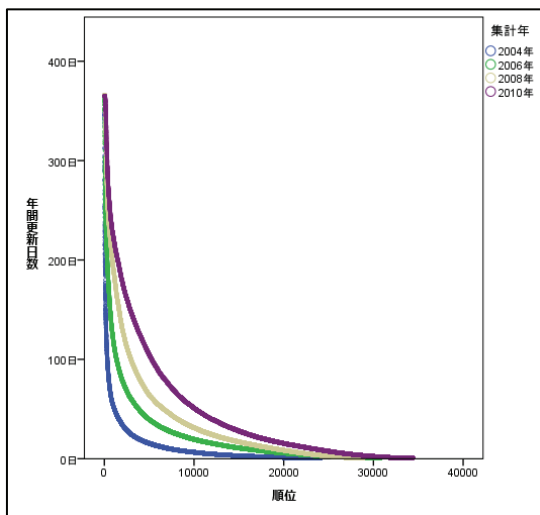


図 3 年間更新回数の年別プロット

この傾向を捉えるため、回帰の曲線推定を用いて、べき乗則 $y = b0[\text{定数}] \times (t^{b1}[\text{係数}])$ への適合を試みた。得られた結果を表 3 に示す。

いずれも決定係数 R^2 が 0.7 以上であり比較的適合度は高いが、経年とともに係数 $b1$ と決定係数は減少する一方、定数 $b0$ が増加する傾向にある。

表 3 変換更新回数分布の曲線推定

年	モデル集計					パラメータ推定値	
	R^2	F	df1	df2	p	定数($b0$)	$b1$
2004	0.88	179520.36	1	24172	***	201052.61	-1.159
2006	0.81	133160.43	1	30839	***	847492.86	-1.216
2008	0.76	89491.38	1	28943	***	1675893.02	-1.245
2010	0.73	90851.71	1	34400	***	4161454.87	-1.303

6. 自治体別集計による経年変化

では、更新履歴を自治体別に集計した場合、どのような傾向が現れるだろうか。

本研究では、集計対象を在籍学校数 6 校以上の 1178 自治体とした。さらに詳細分析として東京 23 区・政令指定都市、2010 年の学校平均年間更新日数の上位 10 自治体、計 51 自治体を抽出した。なお、自治体が運用するサーバ仕様で履歴取得の精度が極端に悪いケースは自治体単位で分析から除外した。

i-learn.jp では学校サイトの更新実績を更新率としてランク表示している。ランク A は学校稼働日を基準とした 60%以上、ランク B

は 48%以上、ランク C は 36%以上、ランク D は 24%以上、ランク E は 12%以上、ランク F は 12%未満である。

この基準を用いて 3 自治体の更新率ランク分布を図 4 に示す。学校総数は A 県 I 市が 61 校、M 区が 32 校、K 県 K 市は 173 校である。

いずれも年を追うごとに高頻度ランク (A ~ B) の分布が増える傾向にあるが、そのスピードには自治体間に大きな差が存在する。

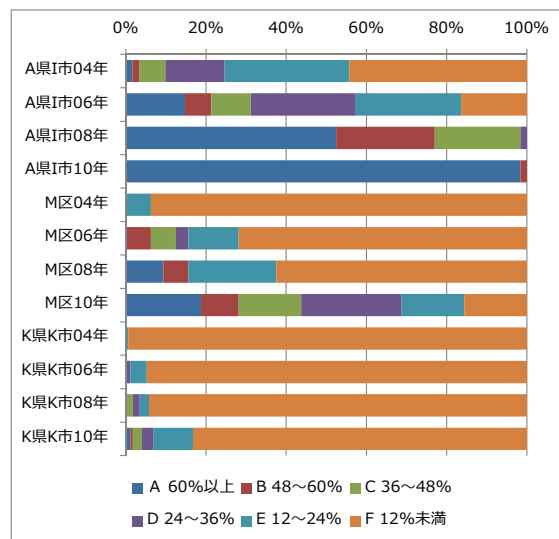


図 4 更新頻度の自治体別ランク推移

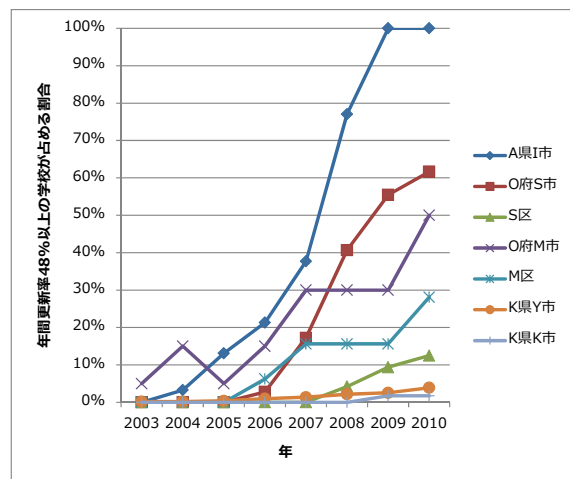


図 5 ランク B 以上を占める割合の推移

ランク B(48%)以上の学校が自治体学校全体に占める割合の年推移を示したものが図 5 である。

100%を達成した A 県 I 市は 2005 ~ 2009 年の 5 年間で、60%強の O 府 S 市は 2007 ~ 2010 年の 4 年間で急速な変化が起こったこ

とが分かるが、一方で、5%前後の領域にずっととどまっている自治体も存在することが分かる。

これら2つのグラフから考察できる事としては、自治体別集計も全体傾向と相似であり、いずれも突出した更新実績を持つ学校が少数存在する構造に変わりはない。

ただし、教育委員会側の何らかの働きかけが効果的に展開される事で、中位に位置する学校の実績が大幅に改善されうることが示唆される。

7. 自治体別集計によるべき乗則の適用

自治体別集計による分布傾向をさらに的確に捉えるため、先の51自治体について回帰の曲線推定を用い、べき乗則 $y = b_0[\text{定数}] \times (t \wedge b_1[\text{係数}])$ への適合を試みた。

曲線推定における決定係数 R^2 は 0.951~0.2、 $b_0[\text{定数}]$ は 210.29~14305.16、 $b_1[\text{係数}]$ は -1.637~0.122 の範囲で分布している。

代表的自治体ケースの結果を学校数・学校の年間平均更新日数・ランク A~B 学校の自治体全体占める割合とともに表4に表す。

さらに、O府S市およびK県K市の年間更新回数分布とべき乗則曲線推定の結果をそれぞれ図6・図7に示した。

年間平均更新日数の高いO府S市は中位に位置する学校の実績が高いので、べき乗則の曲線に適合しないが、年間平均更新日数の低いH県K市はほぼべき乗則の曲線上に観測値がプロットされていることが分かる。

表4 各自治体の曲線推定値等

自治体	学校数	平均更新日数	AB ランク比	b1(係数)	b0(定数)	R ²
A県I市	61	315.39	100.0%	-0.122	459.15	0.66
I県U市	12	225.08	100.0%	-0.529	445.74	0.43
O府S市	145	190.31	80.7%	-0.406	886.87	0.43
A県K市	25	178.96	80.0%	-0.284	381.28	0.80
H市	25	171.64	60.0%	-0.907	1001.68	0.43
M区	32	124.91	43.8%	-0.756	716.53	0.71
N区	38	62.50	15.8%	-0.105	678.34	0.87
H県K市	268	30.79	6.0%	-1.637	14305.16	0.87
K県K市	173	22.50	4.0%	-1.278	2095.89	0.88

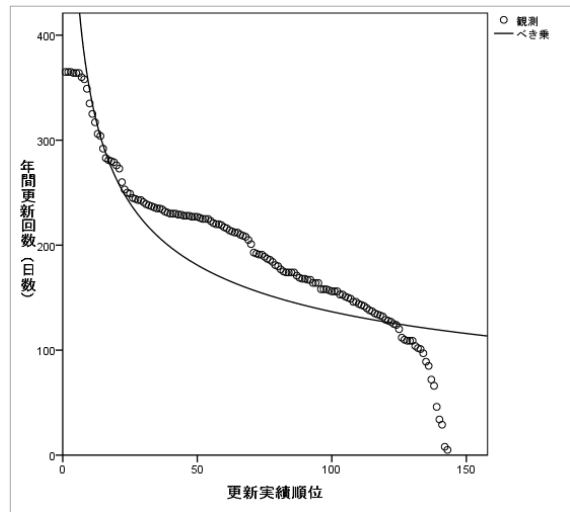


図6 O府S市の年間更新回数分布とべき乗則曲線推定

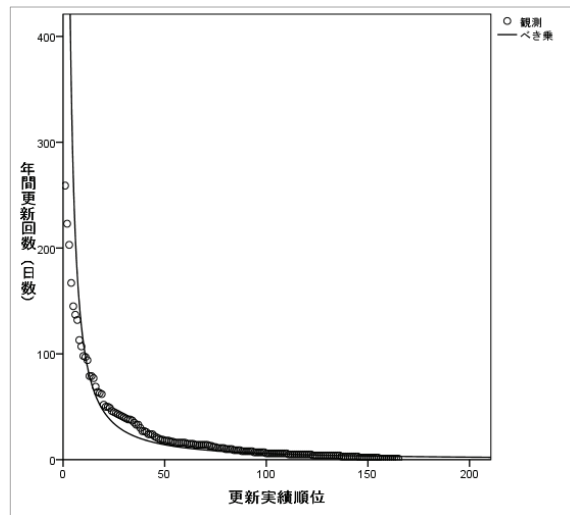


図7 H県K市の年間更新回数分布とべき乗則曲線推定

仮に $b_0[\text{定数}]$ を 100 と固定して、 $b_1[\text{係数}]$ の要素だけを抽出したグラフが図8である。

これによると、 $b_1[\text{係数}]$ の値が低いほど中位の落ち込み方が急になるので、自治体内の学校実績分布としては、 $b_1[\text{係数}]$ が高い(0に近い)、つまり、中位に位置する学校のサイト更新実績を高める事が望ましいといえる。

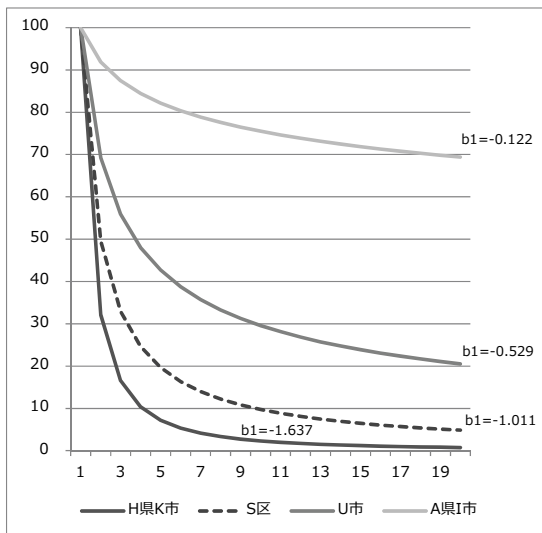


図 8 べき乗則 b_1 [係数]のみの比較

8. 自治体別集計の各指標相関

先に得た b_1 [係数]・ b_0 [定数]と表 4の各要素の関連性を検証するため、Pearson の相関係数 r を求めた。結果を表 5に示す。

表 5 各指標の相関分析

平均更新日数	相関	0.062			
	N	1178			
ABランク比	相関	0.055	0.908		
	N	1178	1178		
b_1 (係数)	相関	-0.308	0.629	0.647	
	N	51	51	51	
b_0 (定数)	相関	0.593	-0.198	-0.191	-0.434
	N	51	51	51	51
	学校数	平均更新日数	ABランク比	b_1 (係数)	

これによると、年間平均更新日数×ABランク比($r=0.908$)、 b_1 [係数]×ABランク比($r=0.649$)、 b_1 [係数]×年間平均更新日数($r=0.629$)、 b_0 [定数]×学校数($r=0.593$)は比較的相関が高いのに対し、学校数×ABランク比($r=0.055$)、学校数×年間平均更新日数($r=0.062$)はほとんど無相関である。

つまり、 b_1 [係数]は年間平均更新日数(ABランク)との関連性が強く、 b_0 [定数]は学校数との関連性が高い。

一方、学校数は年間平均更新日数に影響しないことを示している。

9. 年間平均更新日数×べき乗則係数による自治体プロット

図 9は自治体ごとに求めた年間平均更新日数× b_1 [係数]の散布図である。

これによると、年間平均更新日数が150日以下の自治体は1つを除いて b_1 [係数]-0.75以下である。

一方、年間平均更新日数が150日以上の自治体は b_1 [係数]-0.75以上が多いものの、-0.75以下のケースが3つ存在していることが分かる。

データから直線回帰を求めると、 $R^2=0.396$ 、年間平均更新日数= $138.59 \times b_1$ [係数]+215.75であった。

すなわち、年間平均更新日数の少ない自治体ほど、突出した更新実績を持つ学校が少数存在する構造になりやすいといえる。

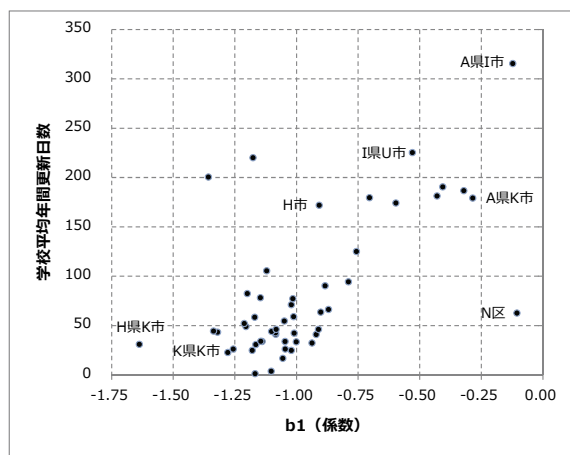


図 9 年間平均更新日数× b_1 [係数]の自治体プロット

10. 考察

本研究の学校サイト更新履歴の統計分析で明らかになった主な事項は次の通りである。

- ・全体レベルの時系列変化として、累積月数と小学校ダミー変数により月間更新回数総和を求める回帰式が得られた。
- ・全体レベルの月変動傾向に一定のパターンが認められ、学校種による違いが明らかになった。
- ・全体レベルの年間更新回数実績順による分布は、およそべき乗則の曲線に従うが、年を

追うごとに適合度が低くなり、将来的には直線や逆S字の分布になることが示唆された。

- ・自治体別集計による経年変化では、いずれも更新実績は上昇傾向にあるが、自治体によってそのスピードには大きな差がある。

- ・自治体別の年間更新日数実績順による分布についてべき乗則の曲線推定を行い、割り出した b_1 [係数]・ b_0 [定数]とその他パラメータとの相関分析を行った。

- ・べき乗則の b_1 [係数]はいずれのケースもマイナス値をとるが、値が低くなるほど中位の落ち込みが急激になる特性を持つ。

- ・べき乗則の b_1 [係数]は年間平均更新日数との関連性が強く、 b_0 [定数]は学校数との関連性が高い。

- ・年間平均更新日数の低い自治体ほどべき乗則への適合度は高く、 b_1 [係数]は低くなる傾向がある、すなわち、突出した更新実績を持つ学校が少数存在する構造になりやすい。

- ・一方、年県平均更新日数の高い自治体ほどべき乗則への適合度は低く、 b_1 [係数]は高くなる、すなわち、中位に位置する学校の実績が向上していることがうかがえる。

そもそも学校サイトの運用は各学校の事情に強く影響される性質をもっているため、学校の自主性に任せた場合は、結果的に突出した更新実績を持つ学校が少数存在する構造になりやすい。本研究の結果（べき乗則への適合度と b_1 [係数]）はそれを裏付けるものである。

また、自治体別集計で年間平均更新日数やランク AB 比が改善した事例では、中位から下位に位置する学校の実績が大幅に改善されていることから、教育委員会側の学校に対するなんらかの働きかけが影響しているものと示唆される。

本研究では、年間平均更新日数(ランク AB 比)の他に、曲線推定によるべき乗則適合度、べき乗則の b_1 [係数]・ b_0 [定数]を指標として用いたが、将来的には、より特性を捉える統計指標をもって自治体取り組みの成果を客観的に評価することが可能であろう。これはデータ駆動型意思決定 (DDDM: Data Driven

Decision Making) への布石となるものである。

ただし、統計指標による客観的評価への課題としては、自治体サーバ仕様による更新履歴取得の精度と、分布の曲線推定における数値の安定性(適合度)があげられる。

特に、曲線推定では、今後べき乗則以外の式をあてはめ、その特性を明らかにする必要があると考えられる。

参考文献

豊福晋平(2002) オンラインデータベースを利用した学校ホームページ群の客観的評価. 日本教育情報学会年会論文集 (18), pp45-46

豊福晋平(2005) 積極的情報開示と外部評価による学校改善. 日本教育工学会研究報告集, JSET05-3, pp149-154

i i-learn.jp 「日本の学校」
<http://www.i-learn.jp/>