

インターネットにおける 検索エンジン利用とインフルエンザの流行との関連

—日本におけるデータを用いた検討—

スエキ ハジメ
末木 新*

目的 わが国におけるインフルエンザに関する検索およびインフルエンザの流行に関する指標を用いて、インフルエンザの流行とインターネット上の検索行動との関連の追試を実施するとともに、検索行動がインフルエンザ流行の先行指標として機能するかを調査する。この際、インフルエンザに関する検索行動をいくつかのクラスタに分割した上で検討する。

方法 インフルエンザに関する検索状況に関するデータは、Google Insights for Searchを利用して収集した。インフルエンザの流行の指標としては、国立感染症研究所感染症情報センターのインフルエンザ様疾患発生報告（学校欠席者数など）を利用した。インフルエンザに関する検索語をクラスタ分析を利用して分類した上で、インフルエンザの流行の指標との相互相関を検討した。

結果 インフルエンザに関する日本語の検索語は大きく「予防」と「対処」の2つに分類されること、「対処」に関する検索がインフルエンザの流行と強い相関（ $r > 0.80$ ）を持つことが示唆された。また、検索を先行させた際の相関（相互相関）は、中程度から強い相関を示していた（ $0.34 < r < 0.85$ ）。

結論 海外のデータで指摘されていたインフルエンザの流行と検索エンジン利用との関連についてわが国のデータを用いて追試を行った結果、先行研究の内容は支持された。「対処」に関する検索の増加は、その後の流行の増加と関連することが示唆されたが、「対処」に関する検索とインフルエンザの流行は同期的に変化していると考えられた。

キーワード インフルエンザ、インターネット、情報疫学、検索エンジン、相互相関

I はじめに

2000年には1割程度であったインターネット利用率も、2009年には8割に迫っており¹⁾、この10年で国民の大半がインターネットを利用するようになった。こうしたインターネットの普及に伴い発展した研究領域として「情報疫学（infodemiology）」が挙げられる。情報疫学とは、インターネットを中心とした電子メディア内に散在する情報をリアルタイムで収集・分析することで地域全体の健康への脅威を扱うアプ

ローチの総称であり²⁾、こうした試みは疾病のみならず自殺などの社会現象をも分析の対象としている³⁾⁴⁾。

この領域の研究は従来、情報の供給側の分析（例：どのようなウェブサイトがどの程度あるか）が中心であったが、ロボット型検索エンジンの利用が普及しその情報が公開されるにつれ、情報の需要側（例：どのような言葉がどれくらい検索されているか）に関する研究³⁾が可能になり、公衆衛生を中心とした専門家の関心を集めている¹⁾。その理由は、検索の結果として表

* 東京大学大学院教育学研究科臨床心理学コース博士課程（日本学術振興会特別研究員）

示される画面の広告を利用することで、問題を抱えている人々を適切な援助資源に結びつけることが可能となるからである。そして、検索エンジンの利用動向を分析した研究としては、Googleの検索データとインフルエンザの流行の関連を指摘した研究⁶⁾が有名である。

季節性のインフルエンザの流行は公衆衛生上の大きな問題であり、世界保健機関によると年に25～50万人の死者が生じている⁷⁾。わが国においてもインフルエンザは重要な公衆衛生上の問題であり、厚生労働省の人口動態調査によると2000年代のインフルエンザによる年次死亡者数は、214～1,818人の間で推移している⁸⁾。流行の早期特定と素早い対策がインフルエンザの影響を軽減することが明らかになっていることを考慮すると⁹⁾¹⁰⁾、Ginsbergら⁶⁾がインターネット上でのインフルエンザに関する検索行動がインフルエンザの流行状況と関連することを見出した知見には重要な意味があると考えられる。

しかしながら、この研究では、①多数のインフルエンザに関する検索語の検索量を一つの独立変数に合成しており、インフルエンザに関する検索動機の多様性が検討されていない、②インフルエンザの流行状況とインフルエンザに関する検索量の関連が検討されたのみであり、検索行動によるインフルエンザ流行の予測可能性が検討されていない、といった問題点が存在する。インターネットを用いたインフルエンザの流行への介入を具体的に考えた場合には、検索連動型広告等を活用してインフルエンザに関する情報を求める利用者に適切な情報提供を実施するといった方策が考えられる。この際、仮に検索行動の状況から流行の予測がたてられていた場合には、自らや周囲の者がインフルエンザに感染してから病院を探すといった対処行動を支援するだけでなく、インフルエンザの流行に先だち適切な予防方法に関する情報提供を行うことでより大きな予防効果を創出することが可能になると考えられる。

そこで本研究では、わが国におけるインフルエンザに関する検索および流行状況に関する指

標を用いて、インフルエンザに関する検索行動をいくつかのクラスタに分割した上でインフルエンザと検索行動の関連の追試を実施するとともに、検索行動がインフルエンザ流行の先行指標として機能するかを検討した。

Ⅱ 方 法

(1) 検索データ

インフルエンザに関する検索状況のデータは、Google Insights for Search (以下、GIS) を利用して収集した。GISとは世界中のGoogleドメインで行われたウェブ検索を分析し、検索された語句やフレーズの検索ボリューム¹¹⁾を出力できるサービスである¹¹⁾。このサービスでは、一定の検索ボリュームがないと結果が出力されない、短期間に同一ユーザーが繰り返し行った同一の検索は結果から除外されるなどデータの信頼性を担保するための工夫がなされている。なお、わが国において利用者の多いYahoo! JAPANの提供するロボット型検索エンジンにおける各検索語の最新の検索状況については、現在一般ユーザーに対して情報が公開されていない。そのため、本調査ではGoogleを利用することとした。

2010年12月に、GISの比較基準を「検索クエリ」、検索クエリを「インフルエンザ」、フィルタを「ウェブ検索」「日本」「すべての地方」「すべてのカテゴリ」、期間を2004年1月から2008年12月とし¹²⁾、「インフルエンザ」に関連の深い人気検索語上位50語を取得した¹³⁾。なお、この50語には「インフルエンザ」自体が含まれていた。さらに、インフルエンザに関する検索語50語の検索ボリュームは、GISの比較基準を「検索クエリ」、検索クエリを「インフルエンザ関連語」(例:「インフルエンザ」「インフルエンザ予防」など)、フィルタを「ウェブ検索」「日本」「すべての地方」「すべてのカテゴリ」、期間を2004年1月から2008年12月とし、結果を出力した。50語の内45語は週単位(2004年1月4日から始まる週～2008年12月28日から始まる週までの261週分)で結果が出力された

表1 各クラスタに含まれる検索語

予防 (17語)	対処 (25語)
インフルエンザ 予防接種, 予防接種 インフルエンザ, 予防, 予防接種, インフルエンザ ワクチン, ワクチン インフルエンザ, インフルエンザ 注射, インフルエンザ 予防注射, インフルエンザ マスク, マスク インフルエンザ, インフルエンザ 対策, 新型インフルエンザ, 新型, 新型インフルエンザ 対策, インフルエンザ 厚生労働省, インフルエンザ 副作用, インフルエンザ 妊娠	インフルエンザ, インフルエンザとは, インフルエンザ 情報, インフルエンザ 感染, インフルエンザ ウイルス, インフルエンザ 薬, インフルエンザ 風邪, 風邪 インフルエンザ, インフルエンザ 今年, 今年 インフルエンザ, インフルエンザ 流行, インフルエンザ 型, インフルエンザ B型, B型 インフルエンザ, インフルエンザ B, A型 インフルエンザ, インフルエンザ A, インフルエンザ 期間, インフルエンザ 潜伏期間, インフルエンザ 症状, インフルエンザ 検査, インフルエンザ タミフル, タミフル インフルエンザ, タミフル, インフルエンザ 脳症

表2 検索ボリュームとインフルエンザの流行指標との相互相関

が、5語（「インフルエンザ A型」「インフルエンザ 潜伏」「インフルエンザの症状」「インフルエンザ ワクチン接種」「インフルエンザ 治療」）は週ごとのデータが出力されるに足る検索ボリュームがなく月ごとに結果が出力されたため、分析から除外した。

時間差 (週)	予防				対処			
	休校数	学年閉鎖学校数	学級閉鎖学校数	患者数	休校数	学年閉鎖学校数	学級閉鎖学校数	患者数
-4	-0.14*	-0.15*	-0.13	-0.14*	0.25**	0.26**	0.31**	0.29**
-3	-0.10	-0.11	-0.10	-0.11	0.41**	0.41**	0.47**	0.44**
-2	-0.04	-0.06	-0.05	-0.06	0.59**	0.59**	0.64**	0.61**
-1	0.03	-0.01	0.00	-0.01	0.75**	0.73**	0.76**	0.74**
0	0.10	0.05	0.05	0.04	0.85**	0.80**	0.81**	0.80**
1	0.15*	0.08	0.09	0.08	0.85**	0.79**	0.78**	0.77**
2	0.15*	0.10	0.10	0.08	0.75**	0.69**	0.67**	0.65**
3	0.15*	0.09	0.09	0.07	0.61**	0.54**	0.50**	0.48**
4	0.15*	0.09	0.09	0.08	0.46**	0.39**	0.35**	0.34**

注 1) *p<0.05, **p<0.01
2) 正の時間差は、検索を先行させた場合であり、負の時間差はその逆を示す

(2) インフルエンザの流行状況

2004年1月4日からの261週分について、国立感染症研究所感染症情報センターのインフルエンザ様疾患発生報告（学校欠席者数）¹²⁾を利用し、休校数、学年閉鎖学校数、学級閉鎖学校数、患者数のデータを作成した。なお、学校での出欠状況はインフルエンザ流行の明確な指標となることが先行研究¹³⁾により指摘されている。統計解析には、SPSS11.0Jを用いた。

III 結 果

分析対象となった45語について、インフルエンザとの関連の強さを調べるため、「インフルエンザ」と残りの44語との相関係数を検討した。その結果、「厚生労働省」「鳥インフルエンザ」「感染症」の3語については弱い相関（r<0.40）しかみられなかったことから、以下の分

析から除外した。

次に、残りの42語の検索ボリュームについて、それぞれ標準化した値を用いてクラスタ分析（平方ユークリッド距離，ward法）を行い、検索語を分類した。その結果、解釈可能な2つのクラスタが抽出された（表1）。1つ目のクラスタは、「予防」「予防接種」「マスク」「対策」といった言葉を含む検索語が多いことから、「予防」とした。2つ目のクラスタは、「情報」「症状」「潜伏期間」「薬」「タミフル」といった言葉を含むことから、「対処」とした。

最後に、検索ボリュームの先行指標としての可能性を検討するために、各クラスタに分類された検索語の検索ボリュームを標準化した値の平均値を独立変数に、インフルエンザ様疾患発生報告（学校欠席者数）に含まれる休校数・学年閉鎖学校数・学級閉鎖学校数・患者数を従属変数とし、それぞれの組み合わせでの相互相関

を検討した(表2)。相互相関のラグ(時差)は1カ月(4週間)とした。その結果、「予防」とインフルエンザの流行の指標との間には統計的に有意な相関もいくつか見られたものの、ほとんど相関は見られなかった($-0.15 < r < 0.15$)。これに対し、「対処」とインフルエンザの流行の指標との間には、時間差なしの時点ではいずれの

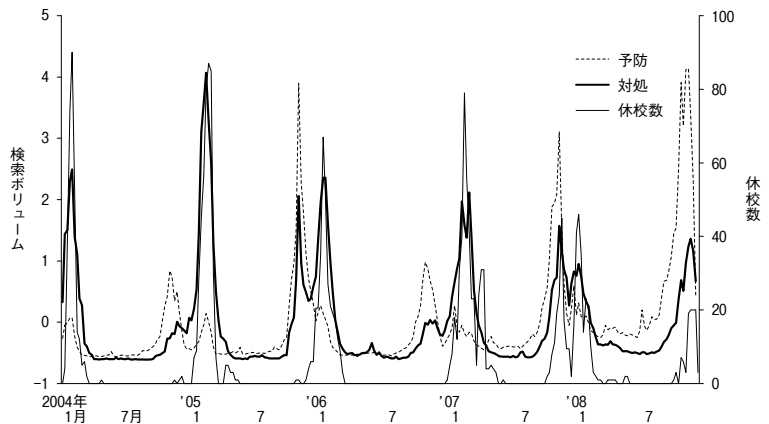
指標との間にも強い相関($r > 0.80$)がみられた。また、検索を先行させた際の相関は、中程度から強い相関を示していた($0.34 < r < 0.85$)。インフルエンザの流行を先行させた際の相関は、検索を先行させた際の相関よりもやや弱く、弱～中程度の相関を示していた($0.25 < r < 0.75$)。

最後に、インフルエンザに関する検索ボリュームと「対処」に関する検索と最も強い関連を示した休校数の時系列的変化を図示した(図1)。

IV 考 察

インフルエンザの流行は特定の季節に生じるものの、その流行の始まる時期を明確につかむことは難しいことであった。市販薬の売れ行き状況に関するデータの利用など、これまでも診断以外の情報を収集することによってインフルエンザの流行の端緒を捉えるための数々の試みがなされてきたが¹⁴⁾¹⁵⁾、本研究ではインターネットの検索エンジン利用という観点からこの問題にアプローチを行った。その結果、インフルエンザに関する日本語の検索語は大きく「予防」と「対処」の2つに分類されること、「対処」に関する検索がインフルエンザの流行と強い関連を持つことが示唆された。

図1 インフルエンザに関する検索ボリュームと休校数の時系列的变化



(1) 検索語の分類とインフルエンザの流行との関連

本研究では各検索語の検索ボリュームを標準化した値を利用してクラスタ分析を実施した。この結果として抽出された2つのクラスタはそれぞれ検索パターンの類似したもの集合である。検索パターンの類似性は検索動機の類似性を示していると考えられるが^{注4)}、それぞれに含まれる検索語の内容から2つのクラスタは「予防」と「対処」とした。

上述のように分類された検索語の内、「予防」に関する検索語とインフルエンザの流行との間には明確な関連は見られなかった。その一方で、「対処」に関する検索語はインフルエンザの流行と強い関連が見られた。図1を参照すると、「予防」に関する検索語は毎年秋頃になると検索ボリュームが上昇するサイクルが見られる。このサイクルは実際のインフルエンザの流行とは関連がないように見受けられる。こうした検索は、予防接種に関する情報探索が中心であるため、インフルエンザの流行以前の話となっている。

そして、「予防」に関するピークの数カ月後にはインフルエンザの流行のピークが訪れ、それと共に「対処」に関する検索ボリュームが増加する。ラグ0での「対処」とインフルエンザの流行は強い相関を示しており($r > 0.80$)、対象期間のいずれの年においてもそのピークは

一致している。「対処」に含まれる検索語には「薬」や「タミフル」といった言葉も含まれるが、こうした検索は検索者本人、あるいは身近な者がインフルエンザに感染してから実施したものだと考えられる。

さらに、ラグ0以外の相関を見ると、「対処」については、検索を先行させた場合においても、また流行を先行させた場合についても統計的に有意な正の相関を示していた。検索を先行させた際の相関 ($0.34 < r < 0.85$) が流行を先行させた際の相関 ($0.25 < r < 0.75$) よりもやや強い値を示していることも考慮すると、「対処」に関する検索行動の増加はその後のインフルエンザの流行の先行指標として機能しうると考えられる。しかしながら、統計的な分析からは先行指標としての機能についても言及可能であるものの、図1の時系列的な変化からは先行しているというよりは、ほぼ同期（同じく）して変化していると考えの方がより妥当な解釈であると思われる。

(2) 本研究の限界と今後の展望

最後に、本研究の限界と今後の展望を述べる。本研究の限界の1点目は、2004年から2008年までの5年間のデータのみを利用していることである。時系列的な関連の検討にはより長期のデータを扱った研究を実施する必要があると思われる。インターネットにおけるロボット型検索エンジンの利用の一般化は現代的な現象であり、より多くのデータが集積されてからの検討が望まれる。2点目は、Googleの検索データのみを用いている点である。わが国での検索エンジンの利用は、Googleが4割、Yahoo! JAPANが5割であり¹⁶⁾、利用の半数を占めるYahoo! JAPANのデータを扱うことはできていない。

以上のような限界はあるものの、本研究ではこれまでに指摘されていたインフルエンザの流行と検索エンジン利用の関連についてわが国のデータを用いて追試を行った点、検索エンジン利用の先行指標としての可能性を検討した点に一定の意義があると考えられる。公衆衛生へのインターネット関連技術の活用は発展段階にあ

り、より多様な工夫の余地がある。本研究では情報の需要側の研究として検索エンジン利用を取り上げたが、今後はインフルエンザ等の疾病に関する情報を提供するサイトの中でも多数のアクセスを集めるもの（例：国立感染症研究所感染症情報センター、<http://idsc.nih.gov/index-j.html>）のアクセス解析等の研究を合わせて用いることで、より詳細な流行の分析が可能になると考えられる。

- 注1) GISでは検索回数ではなく、指定した期間・地域における特定の検索語の相対的検索頻度が、最高値が100になるよう正規化されて出力される¹¹⁾。そのため、本論文では「検索回数」ではなく「検索ボリューム」という用語を用いる。
- 2) GISのデータ提供期間は2004年以降である¹¹⁾。また、2009年には夏頃から流行した新型インフルエンザにより爆発的に検索数が増加したため、相対的検索ボリュームのみしか出力しないGISでこの時期を対象期間に含めてデータを取得すると他の時期の検索ボリュームの標準偏差が小さくなりすぎる。そこで、2009年のデータは除外し、対象期間を2008年12月までとした。
- 3) GISでは検索行動の時系列的前後関係を利用して検索語間の関連を算出している¹¹⁾。
- 4) たとえば、個々の利用者の行動という観点から考えると、「インフルエンザ」という語で求める情報が得られない場合、同一の動機で類似の検索語（例：「インフルエンザとは」）を用いた検索を続けて行う。そのため、同一の動機のもとで利用される検索語の検索ボリュームは同じ時期に高まりやすく、時系列的な検索パターンも似た形をとりやすくなると考えられる。

文 献

- 1) 総務省. 通信利用動向調査（平成21年）. (<http://www.soumu.go.jp/johotsusintokei/statistics/statistics05.html>) 2010.12.20.
- 2) Eysenbach G. Infodemiology and infoveillance: Framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet.

- Journal of Medical Internet Research 2009 ; 11 (1) : e11.
- 3) McCarthy MJ. Internet monitoring of suicide risk in the population. *Journal of Affective Disorders* 2010 ; 122(3) : 277-9.
 - 4) 末木新. インターネットにおける自殺関連語の検索ボリュームと既遂自殺の関連 - Google Insights for Search betaを用いた検討. *臨床心理学* 2011 ; 11(1) : 77-82.
 - 5) Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection: Harnessing the web for public health surveillance. *The New England Journal of Medicine* 2009 ; 360(21) : 2153-7.
 - 6) Ginsberg J, Mohebbi MH, Patel RS, et al. Detecting influenza epidemics using search engine query data. *Nature* 2009 ; 457(7232) : 1012-4.
 - 7) World Health Organization. Influenza fact sheet. (<http://www.who.int/mediacentre/factsheets/fs211/en/>) 2010.12.20.
 - 8) 厚生労働省. 人口動態調査 (http://www.e-stat.go.jp/SG1/estat/GL08020101.do?_toGL08020101_&tstatCode=000001028897&requestSender=dsearch) 2010.12.20.
 - 9) Ferguson NM, Cummings DAT, Cauchemez, S, et al. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* 2005 ; 437 : 209-14.
 - 10) Longini IM, Nizam A, Xu, S, et al. Containing pandemic influenza at the source. *Science* 2005 ; 309 (5737) : 1083-7.
 - 11) Google. Google Insights for Search ほんどのように使えますか? (<http://www.google.com/support/insights/bin/topic.py?topic=13761>) 2010.2.28.
 - 12) 国立感染症研究所感染症情報センター. インフルエンザ様疾患発生報告 (学校欠席者数). (<http://idsc.nih.go.jp/idwr/kanja/infreport/report.html>) 2010.12.10.
 - 13) Lenaway DD, Ambler A. Evaluation of a school-based influenza surveillance system. *Public Health Reports* 1995 ; 110(3) : 333-7.
 - 14) Hogan WR, Tsui FC, Ivanov O, et al. Indiana-Pennsylvania-Utah Collaboration: Detection of pediatric respiratory and diarrheal outbreaks from sales of over-the-counter electrolyte products. *The Journal of the American Medical Informatics Association* 2003 ; 10(6) : 555-62.
 - 15) Davies GR, Finch RG. Sales of over-the-counter remedies as an early warning system for winter bed crises. *Clinical Microbiology and Infection* 2003 ; 9 (8) : 858-63.
 - 16) comScore. Yahoo! attracts more than half of all searches conducted in Japan in January 2009. (http://www.comscore.com/Press_Events/Press_Releases/2009/3/Japan_Search_Engine_Rankings) 2010.3.19.