

英文経済雑誌記事内容把握のための理論とソフトウェアツール

○平田周、横山誠

A Theory and a Computer Tool for Quick Understanding of an Economic Article

○Schu Hirata, Makoto Yokoyama

While search technologies have made marked progress in recent years, it is yet time consuming and laborious to read references output after searching. Particularly so, to Japanese reading English journals. From our experiences, we knew that sentences having numbers are more meaningful in an article. We came to an idea that if such number-containing sentences are extracted, it will be helpful for a reader to quickly grasp an outline of the article. This report is a result of our examination of the theory, and we developed a PC tool for extracting such number-containing sentences.

1. はじめに

情報検索の理論やツールは進んでいるが、一つの論文や雑誌記事、とりわけそれが英文の場合、すばやくその要点を把握するという方法は、速読法などが提唱されているものの、ソフトウェアツールやそのベースになる理論はない。英文経済記事の場合、数字が入っている文に重要な意味を含んでいるという経験的仮説をもとに、数字的な語を含む文章を抜き出す PC ツールを開発して、その妥当性を検証してみた。

2. 問題の所在

情報検索技術は、サーチエンジンをはじめ、マークアップ言語など、関心も高く、技術的にも進歩がみられる。これにより、無数の情報から必要な文献や記事が容易に取り出せるようになった。しかし、必要な文献が入手できたとしても、それを読むために要する時間や手間に関しては、旧態依然である。検索結果により提示される文献や記事の数がますます増えていけば、従来にも増して、読み手としての労力は増大せざるをえない。

熟練した研究者であれば、自分が知りたいこと、関心ある問題が明確に意識あるいは正確に定義されているだけに、必要な箇所をすばやく見つけ出すことはさほど苦勞ではないかもしれないが、読む目的が一般的な場合、文献や記事から有用な意味を短時間に読み取することはそれほど簡単ではない。

世に「速読法」なるものがいろいろと紹介されている。眼球を動かさずに文字を追っていくもの、斜め読みするもの、キーワードを読み取るものなどさまざまである。日本語で書かれている場合、そうした方法は、ある程度効果は認められる。しかし、これが外国語となれば、表意文字の日本語と異なり、これらの方法はあまり通用しないように思われる。英語にも速読法はあるが、外国語として読まねばならない私たちには無理もある。相当に英語に精通している人でも、読みとばしていくのはなかなか難しいものである。

速読法に対する疑問は、早く読めるようになることと、理解とは必ずしも一致しないことである。小説やエッセーのようなものであれば、大筋をつかむために、早や読みも効果がなくはないが、意味深い箇所をじっくり理解することが犠牲にされがちである。まして

や専門的な内容であれば、なおさらにそうである。

検索技術は大いに進歩したが、内容理解についての方法論はまだほとんど手つかずの研究領域とっていいであろう。英文で書かれた経済誌の記事を毎週数多く読み、日本語で要約をするという仕事をしていた関係で、ある記事のなかの価値ある情報をすばやく知ることが求められていた。その一つの要領として発見したのは、数字が含まれる文章が、その記事のなかの有意情報になるということであった。

以下に、その妥当性を検証してみた。また、そのために、数字に類する語が含まれる文を自動的に選択するコンピュータ・プログラムを開発した。

3. 数字を含む文章からの内容推定

代表的な英文経済誌 BusinessWeek の記事を例にもちいて、数字に類する語を含む文章だけで、どれだけ内容を理解できるかを検証してみたい。

検証に使った記事は、BusinessWeek 誌 2003 年 4 月 28 日号の「BREAKDOWN」で、副題は How China's decentralized health care is failing となっている。雑誌のタイトルは、その記事の内容を知る最初の手がかりになる可能性は高いが、読者の興味を引くことを目的としている場合も多く、それだけで内容を理解することは一般に難しい。

まずこの記事から、数字的な語を含む文章を選び出してみよう。

"I am not allowed to talk about my country's public health," the physician, from **one** of Beijing's premier hospitals, said to a reporter arriving for a scheduled interview.

約束したインタビューを行ったレポーターに、北京のある大手病院の医師は、「私は、わが国の公衆衛生について話すことを許されていない」と語った。

China's vast network of public hospitals, rural clinics, and private facilities is now so inequitable and inefficient that many public health professionals see the country as **one** of the weakest links in the global medical system.

中国の公立病院、地方の診療所、私的医療機関の広範囲のネットワークは、不均等で非効率的であることから、公衆衛生の専門家の多くが世界の医療システムのなかでも最も弱いつながりの国の 1 つだとみている。

Already more than **5 million** Chinese suffer from tuberculosis, the second-highest number on the planet - and a large number of the cases are drug-resistant.

すでに 500 万人以上の中国人が結核を患っており、これは地球上で第 2 番目に多い。しかも、ほとんどのケースで、薬剤耐性になっている。

China is also belatedly waking up to a crisis with AIDS, already estimated to affect more than **1 million** people.

遅きにすぎたとはいえ、中国はエイズの危機を意識しているが、感染者は 100 万人を超すと推定される。

That number could soar to **10 million** to **20 million** by **2010**, making China one of the hardest hit countries in the world, says a recent report by the Center for Strategic &

International Studies in Washington.

2010 年までに、この数は 1,000 万人、2,000 万人へと膨らむ可能性があり、中国が世界で最もひどいエイズ国の一つになってしまうと、ワシントンにある Center for Strategic & International Studies は最近報告している。

The spread of the disease from southern China looks to be producing the **first** global epidemic in the **21st** century.

中国南部から広がったこの病気は、21 世紀最初の地球規模での流行病となりつつあるようにみえる。

Sources say Vice-Minister Wu Yi has just pledged **\$55 million** to fight SARS.

その筋によれば、Wu Yi 副首相は、SARS 対策費として 5,500 万ドルを約束した。

Premier Wen Jiabao, on April **13**, said that although progress against SARS had been made, "the overall situation remains grave,"

4 月 13 日、Wen Jiabao 首相は、SARS への対策は進んだが、全体的な状況は未だに深刻だと述べた。

With relatively low expenditures, the mainland was able to ensure basic health for **85%** of its population, slashing infant and maternal mortality rates.

比較的少ない費用で、中国本土全人口の 85%の基本的な保健を維持でき、乳幼児と母親の死亡率を引き下げた。

"China used to be regarded as **one** of the great success stories" in public health, says Jagadish Upadhyay, a lead project officer for health in China at the World Bank in Washington.

ワシントンの世界銀行で中国の保健問題を担当するプロジェクト責任者 Jagadish Upadhyay 氏は、「中国はかつて、最も成功した例の一つとみなされていた」と語る。

That quickly began to change, however, when Deng Xiaoping launched economic reforms in **1979** and carried out a massive decentralization that devolved responsibility and funding of public health to the provinces, cities, and even townships.

しかし、ひじょうに早い速度で変化が始まったのは、1979 年に Deng Xiaoping 氏が経済改革を始め、大規模な地方分権化により公衆保健に関する責任や予算が、州や都市、さらには町区にまでまかされるようになった時である。

As a result, the portion of total health care funded by Beijing has fallen from **36%** in **1980** to less than **20%** now.

その結果、北京政府の保健予算は 1980 年の 36%から、現在は 20%以下にまで落ちている。

After all, China's total health expenditures have soared more than **tenfold** in that

period, to **\$48 billion** today.

けっきょく、中国の保健関連支出の総額は、その期間に 10 倍以上となり、現在では 480 億ドルに増えている。

The **seven** provinces and autonomous regions in the far west - a vast area home to China's poorest people - only get **5%**.

西部に広がる中国で最も貧しい人々が住む広大な面積の 7 つの省と自治区は、わずか 5% しか受けていない。

Health-care costs have jumped **500%** to **600%** over the last **10** years.

過去 10 年で、保健関連費用は 5 倍にも 6 倍にもはね上がった。

Not only is the care available markedly worse for China's **800 million** rural inhabitants and migrant workers but the vast majority of the rural population, some **90%**, are without any insurance at all.

8 億人にのぼる中国農村部の住民および移住労働者が、きわめて粗悪な保健医療しか受けられないだけでなく、彼らの大多数、約 9 割は医療保険にまったく入っていない。

Urban residents are much better served under a medical social insurance program initiated in **1998** by former Premier Zhu Rongji or through other programs.

1998 年に Zhu Rongji 元首相が始めた医療社会保険制度あるいはその他の計画により、都市部の住民はずっと内容のよいサービスを受けている。

But a full **45%** who are self-employed, jobless, elderly, or work for private companies are uninsured.

しかし、45%を占める自営業、失業者、高齢者、私企業の職場の人たちは、保険に入っていない。

Take **one 23**-year-old Beijing tour guide, whose business has been badly hit by the SARS scare.

北京でツアガイドをしている 23 歳の一人は、SARS の恐怖によって仕事は深刻な打撃を受けた。

以上、数値的な語を含む文章だけを訳出したが、これを一つの文章にしてみた。

約束したインタビューを行ったレポーターに、北京のある大手病院の医師は、「私は、わが国の公衆衛生について話すことを許されていない」と語った。中国の公立病院、地方の診療所、私的医療機関の広範囲のネットワークは、不均等で非効率的であることから、公衆衛生の専門家の多くが世界の医療システムのなかでも最も弱いつながりの国の一つだとみている。すでに 500 万人以上の中国人が結核を患っており、これは地球上で第 2 番目に多い。しかも、ほとんどのケースで、薬剤耐性になっている。遅きにすぎたとはいえ、中国はエイズの危機を意識しているが、感染者は 100 万人を超すと推定される。2010 年ま

で、この数は 1,000 万人、2,000 万人へと膨らむ可能性があり、中国が世界で最もひどいエイズ国の一つになってしまうと、ワシントンの Center for Strategic & International Studies は最近報告している。

中国南部から広がったこの病気は、21 世紀最初の地球規模での流行病となりつつあるようにみえる。その筋によれば、Wu Yi 副首相は、SARS 対策費として 5,500 万ドルを約束した。4 月 13 日、Wen Jiabao 首相は、SARS への対策は進んでいるが、全体的な状況は未だに深刻だと述べた。

比較的少ない費用で、中国本土の全人口の 85% の基本的な保健を維持でき、乳幼児および母親の死亡率を引き下げた。ワシントンの世界銀行で中国の保健問題を担当するプロジェクト責任者 Jagadish Upadhyay 氏は、「中国はかつて最も成功した例の一つとみなされていた」と語る。しかし、ひじょうに早い速度で変化が始まったのは、1979 年に Deng Xiaoping 氏が経済改革を始め、大規模な地方分権化により公衆保健に関する責任や予算が、州や都市、さらに町区にまでまかされるようになった時である。その結果、北京政府の保健予算は 1980 年の 36% から、現在は 20% 以下にまで落ちている。けっきょく、中国の保健関連支出の総額は、その期間に 10 倍以上となり、現在では 480 億ドルに増えている。西部に広がる中国で最も貧しい人々が住む広大な面積の 7 つの省と自治区は、わずか 5% しか受けていない。過去 10 年で、保健関連費用は 5 倍にも 6 倍にもはね上がった。8 億人にのぼる中国農村部の住民や移住労働者は、きわめて粗悪な保健医療しか受けられないだけでなく、彼らの大多数、約 9 割は医療保険にまったく入っていない。1998 年に Zhu Rongji 元首相が始めた医療社会保険制度あるいはその他の計画により、都市部の住人はずっと内容のよいサービスを受けている。しかし、45% を占める自営業、失業者、高齢者、私企業の職場の人たちは、保険に入っていない。北京でツアガイドをしている 23 歳の一人は、SARS の恐怖によって仕事は深刻な打撃を受けた。

以上からもわかるように、この抜粋訳で全体が語ろうとしていることの内容はほぼつかめるであろう。この記事の本文全体の文数は 59 語で、数値的な語を含む文が 19 だから、その比率は 32.2% である。これはかなりの労力削減となろう。

BusinessWeek の他の記事について、全体の文数と数値的な語を含む文の比率をみると、ほぼ 25% から 35% である。イギリスの経済誌 The Economist でもほぼ同様の水準を示す。

数値的な語を含む文を選び出すのに効率的な PC ツールを開発した。数値を含む文章を全体の中から識別し、その文章にアンダーラインを引き、数値に影や色をつける。アンダーラインが引かれた文章だけを取り出すといったことを、PC に作業させることができれば便利である。文章は、. ? ! : . " のいずれか 2 つによって挟まれた文字列と定義した。また数値は、1 2 3 4 5 6 7 8 9 0 のほかに、one two three four five six seven eight nine ten zero、first second third、hundred thousand million billion trillion を加えた。

4. まとめ

上記のような数値的な語を含む文章群 (粗訳) をつくり、これにより記事の内容をかなり具体的に理解できること明らかにした。この粗訳から要約文を作成することが可能である。次に、完成した要約文を示す。

「中国では、500 万人を超す結核患者がおり、ウイルスの薬剤耐性のため薬も効かなく

なっている。結核だけではない。エイズの感染者は 100 万人を超すと推定され、2010 年には、1,000 万人、2,000 万人へと膨らむ可能性がある、ワシントンの戦略国際研究センターは報告している。

中国南部に発生した SARS は、世界に蔓延し、多くの被害を出している。呉儀副首相は、SARS 対策費として 5,500 万ドルを約束したとされ、4 月 13 日、温家宝首相は、SARS への対策は進んでいるが、全体的な状況は未だに深刻だと述べた。

中国では、かつて本土の全人口の 85%に基本的な保健管理がなされ、「中国は以前、保健衛生で最も成功した例の一つとみなされていた」と世界銀行の中国保健問題プロジェクト責任者 Jagadish Upadhyay 氏は語る。その体制が崩れはじめたのは、1979 年に始まった鄧小平の経済改革以降で、大規模な地方分権化により公衆衛生に関する責任や予算が州や都市、あるいは町区に委譲されことに原因がある。その結果、北京政府の保健予算は大幅に減ったが、分配の面で地域的不公平さを生む結果となった。中国で最も貧しい西部の 7 つの省と自治区が受ける保健補助金は全体の 5%でしかない。

それでも中国政府の保健関連支出の総額は膨らみ、現在は 480 億ドルにもなっている。過去 10 年で、保健関連費用も 5 倍にも 6 倍にもはね上がった。しかし、8 億人にのぼる中国農村部の住人らはきわめて粗悪な保健医療しか受けられないうえ、約 9 割の住民は医療保険に入っていないのが実情である。その点、都市部の住人は、医療社会保険制度もあり、ずっと恵まれている。それでも、45%を占める自営業、失業者、高齢者、私企業で働く人ちは保険に入っていない。

今回の SARS 騒ぎで、はからずも中国の医療保健制度の現状を世界に知らせる結果となったといえよう。」

まだ方法論としてはプリミティブではあるが、有意の箇所をキーワードやタグによって検索するやり方ではなく、セマンティックス的なアプローチを試みたことに、多少の意義が認められるならば幸いである。

平田 周 立正大学大学院経営学研究科 LDG01371@nifty.ne.jp

横山 誠 エスブレイン代表取締役 yokoyama@esbrain.com