ビッグデータを活用

本誌編集部が景気指数づくりに挑戦
筑波大、NTTデータと「つぶやき」解析

ツイッターのつぶやきから景気動向指数を作成することは可能か。筑波大学の山本幹雄研究室、NTTデータの協力を得て、編集部がビッグデータ活用に挑戦。「エコノミスト景気指数」をはじいてみた。

安倍政権の経済政策「アベノミクス」により急激な円安・株高が進む。実体経済を潤し、賃金アップまで行き着けるのか、それとも腰折れするのか。2年後の物価上昇率2%を掲げて、日本銀行の黒田東彦総裁が異次元緩和に踏み切ったこともあり、景気や物価を示す統計データへの関心が高まっている。

しかし、月次で発表される内閣府の景気動向指数（CI、一致指数）も、消費者物価指数（CPI）も公表は翌月末以降だ。より早くその値を知るために、言語処理に詳しい筑波大学の山本幹雄研究室と、ツイッターの日本語もしくは日本国内でつぶやかれた全データの取得権を持つNTTデータの協力を得て、つぶやきから経済指標を予測するモデル作りに挑戦した。

まず処理の効率化のため、ツイッターのつぶやきのうち、10年4月～12年12月の5%無作為抽出データを使って、CIとの相関が高いキーワード300語を抽出した。これらには1カ月当たりの出現数が2000回以上などの条件をつけた。

キーワード29語で予測式

次にNTTデータが、選び出した300語について、08年1月～12年3月のツイッターの全データにおけるそれぞれの出現数をカウント。そのデータを基に山本教授らが重回帰分析の複数の手法を試し、数万個の予測式の候補を作成した。予測式はいずれも、ツイッター上で使用された数十程度の言葉について、それぞれの1カ月当たりの出現率に、定数を掛けた総和で表されている。テスト区間とした12年4月～13年3月の出現率をあてはめて、この間の景気指標を予測。予測値と内閣府発表のCIとの誤差（平方和）が最も小さい予測式を採用し、時系列グラフを図に示した。なお、予測式に使用したキーワードは、ここでは公表できない。公表による予測式への影響を避けるためだ。

この予測式はキーワード29語を使い、CIの増減比を予測したモデル。アベノミクスで株価上昇が始まる12年11月以降を、それ以前のつぶやきから生成した予測式でどこまで予測できるのか注目されたが、13年3月まではおおむね似通ったカーブを描いている。

ツイッターには営利目的の大量つぶやき「スパム」が含まれるため、

ツイッターのつぶやきでCIを予測する

（出所）CIは内閣府。予測値は山本幹雄研究室

CIを予測するモデルを作成した大学院生の角田孝昭さん（左）、山本教授（中央）、大学生の中島光夫さん

予測式作りが攪乱されやすいという事情もあるなか、「エコノミスト景気指標」は産声を上げた。

さて結果は？

山本教授は「ツイッターでどこまでできるだろうかという気持ちで臨んだ。評価は未来（4月分のCI公表）に任せるが、手応えを感じた。他の経済指標でも挑戦してみたい」。ただ、誤差が最小なため今回採用した予測式は、直前のCIの値まで極力近づくように作り込んであるだけに「予測するという目的においては、テスト期間のデータを全く使わずに組み立てた予測式で誤差を最小にできればよかった」と話している。

一方、テスト期間のデータを使わずに生成した別の予測式では、アベノミクスで株価が上昇し始めた昨年末以降、CIとの乖離がみられた。300語の絞り込み期間が東日本大震災の前後に重なり、300語には震災関連の言葉が多く含まれていた。ところが回帰式作成区間にはさらにリーマン・ショックがあり、300語の多くが無効になったことが影響した可能性があるという。

今回採用した予測式で4月のCIを予測したところ、その値は91・6。3月のCI値（改定値）や3月のエコノミスト景気指数の予測値よりも下回るが、さて4月のCIはいかに。内閣府によると、過去の景気拡張局面でもCIが単月で低下する不規則な動きはあったという。内閣府は速報値を6月7日に公表する。

ヤフーも景気指標公表

ヤフーも自社の検索キーワードのビッグデータを使い、CIの近似値を算出する「Yahoo！ JAPAN景気指数」を公表している。検索で使われた75億語を60万語に絞り込み、CIと相関の高い196語で予測式を作成した。4月半ばに発表し、3月のCIを91・9と予測。内閣府の公表値は改定値で93・8だった。

担当したヤフーの安宅和人事業戦略統括本部長は「景気指数は継続して公表したい。政治、経済、ヘルスケアは、ビッグデータの3大領域だと思う。様々な可能性を試したい」と話している。

なお、ヤフーも本誌同様、予測式に使用したキーワードは公表していない。

（編集部）

東大が日次物価指数を公表開始 もう一つの「体温計」目指す

販売時点情報管理（POS）データから物価指数を算出する取り組みも始まった。

東京大学の渡辺努教授は5月20日、独自の消費者物価指数「東大日次物価指数（以下、東大指数）」の公表を始めた。全国約300のスーパーの販売価格を集めたPOSデータから算出。日次データを原則5日後に東大ホームページで公表する。日銀が2%の物価目標を掲げる中、総務省の消費者物価指数（CPI）と併せて注目を集めそうだ。

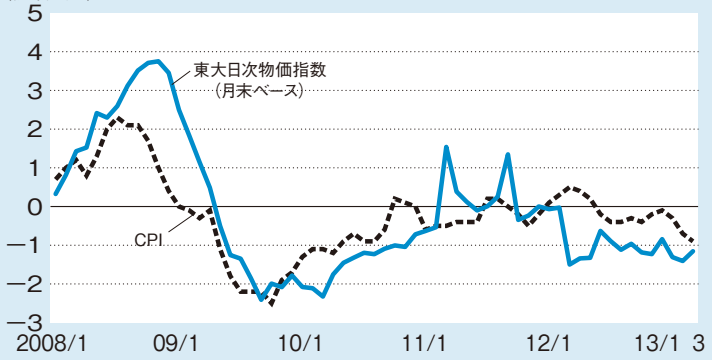
東大指数は、生鮮食料品を除く食品や日用雑貨約20万点の実売価格に、販売数量を加味したウエート付けをして日々計算する。CPIに含まれているデジタル家電などの耐久消費財や家賃などは含まれていないが、POSデータを基にするため、消費者の商品選択が直接反映されるのが特徴だ。スイスやスウェーデンでは、政府がPOSデータで公式CPIを算出している。

一方、総務省のCPIは、特定銘柄で構成され、品目の変更に一定の時間がかかる。指数を計算するための品目ごとのウエートの見直しも原則5年に一度とスパンが長い。東大指数は、こうしたCPIの課題をある程度解消した形になっている。両者のグラフを並べてみると、これまで指摘されてきたCPIの上方バイアスも見て取れる。

物価は経済の体温だ。渡辺教授は「CPIは専門家が作った信頼ある値。公式は公式としつつ、『デフレは終わった』と判断するときに、もう一つの体温計として参考になるといい」と話している。

（編集部）

消費者の実感に近い物価動向をつかめる東大日次物価指数



（出所）東京大学・渡辺努教授、CPIは総務省