

令和3年2月4日

景気ウォッチャー調査を用いたテキスト分析の方法について（第2回）

茨城県政策企画部統計課 企画分析グループ

1 はじめに

人々の景況感を掴むことを目的に、茨城県統計課（以下「当課」という。）が実施している「茨城県景気ウォッチャー調査」（以下「本調査」という。）は、約300名の調査客体の景気判断理由をテキストベースで得られるという特徴がある。このような数値化できないデータを視覚的に分かりやすくして俯瞰できるようにするテキスト分析の方法として、前回²の記事では、単語間の共起（ある2つの語が同じ文章中出现すること。）関係を可視化する「共起ネットワーク図」を中心に紹介し、調査客体の景気判断理由を考察した。しかし、そこで紹介した方法は、単月の結果分析に留まっており、複数月を一度に分析するには、別の方法を用い

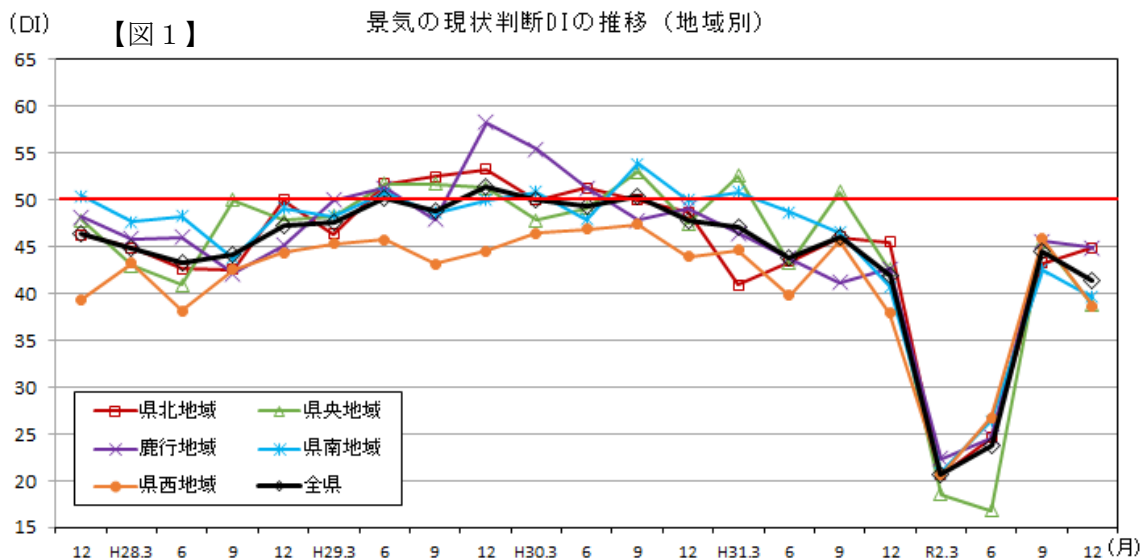
る必要がある。

そこで今回は、複数の条件を指定して、あるキーワードの出現頻度を見られる「コーディングルール（条件指定）」という方法を紹介し、時系列で景気判断理由の特徴を俯瞰していく。

なお、分析にあたっては、前回と同様に「KH Coder」³を用いる。また、分析に用いるデータ⁴は当課公式サイトで公表している。

2 最近の動向と今回の分析対象期間

本調査の景気の現状判断DI（調査時点の景況感を示す指標。以下「DI」という。）の推移を見ると、令和2年3月調査のDI値が大きく低下し（図1）、DI値が本調査開始以



¹ 調査概要：<https://www.pref.ibaraki.jp/kikaku/tokei/fukyu/tokei/betsu/bukka/watch/chogai.html>

² 前回の記事：<https://www.pref.ibaraki.jp/kikaku/tokei/fukyu/tokei/betsu/bukka/watch/bunseki/index.html>

³ テキストデータを統計的に分析するためのフリーソフトウェア。（<http://kncoder.net> よりダウンロード可能）

⁴ 分析に用いるデータ：<https://www.pref.ibaraki.jp/kikaku/tokei/fukyu/tokei/betsu/bukka/watch/bunseki/index.html>

来過去最低となった平成 21 年 3 月の DI 値⁵と同程度にまで落ち込んだ。

今回の時系列分析では、リーマン・ショックが起こった時期から新型コロナウイルス拡大期（ここでは、令和 2 年 3 月から 12 月までの期間とする。）までの期間の様子を見ることとし、リーマン・ショックが発生した月（平成 20 年 9 月）から直近の調査（令和 2 年 12 月）までの 50 期分の調査結果を分析の対象として見ていく。

3 コーディングルールの設定

各調査結果で、どのようなキーワードが何件のコメントで見られるかを集計する方法として、身近なものに Excel の検索機能があるが、今回の分析ではコーディングルール・ファイルを KH Coder に読み込ませて集計する。

コーディングルール・ファイルとは、ある注目したいキーワードを検索するための条件（コーディングルール）を記述したテキストファイルを指す。ファイルには、アスタリスク「*」を先頭に付して注目したいキーワードを記述し、次の行以降に検索の条件を設定する。

*注目したいキーワードの名前 条件式

例えば、収入の減少について述べられているコメントを集計したいとする。この場合、単に「収入」だけで検索すると、収入が上がったというコメントもカウントされてしまう。また、「収入減少」で検索しても、

次のような場合は集計から漏れてしまう。

- ・ 検索対象の一部が離れている場合
「今月は昨年同月よりも収入が減少した。」
- ・ 同義語や類義語で表現されている場合
「最近給料がカットされた。」

このような集計漏れを回避するには、コメントを全て確認するか複雑な条件を指定して検索する必要がある。上の例の場合、コーディングルール・ファイルに次のように記述することで対応できる。

*収入減少 (収入 or 給料 or 給与 or ボーナス) and (減る or 減少 or 下がる or カット)

注目したいキーワードの名前は「収入減少」とし、2 行目以降の条件式は次のように解釈する。

コメントの中に、 「収入」、「給料」、「給与」、「ボーナス」 のいずれかの語を含み、かつ、 「減る」、「減少」、「下がる」、「カット」 のいずれかの語を含んでいる。
--

以上の条件を設定すれば、先述のような集計漏れを減らすことができる。

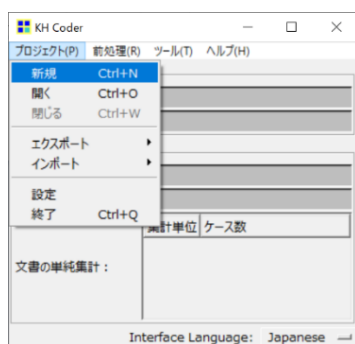
なお、今回の分析で用いるコーディングルール・ファイルについても、当課公式サイトで公表している。

4 分析の流れ

(1) 新規プロジェクトの作成

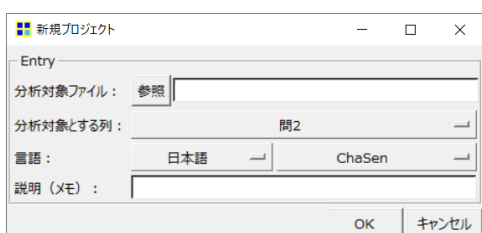
KH Coder を起動し、[プロジェクト (P) - 新規] から新規プロジェクトを作成する。

⁵ リーマン・ショック期は DI 値が大きく低下し、平成 21 年 3 月には本調査開始以来過去最低の 18.5 となった。



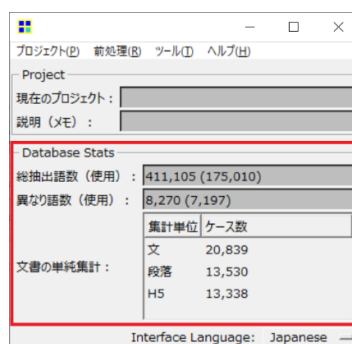
(2) 分析対象ファイルの読み込み

「参照」ボタンから、平成 20 年 9 月から令和 2 年 12 月までの景気の現状判断理由のテキストデータをまとめた Excel ファイル「2008.09_2020.12 茨城県景気ウォッチャー調査現状判断理由.xlsx」を選択して読み込む。この際、分析対象とする列を指定する。ここでは、現状の景気判断理由が書かれている「問 2」列を選択する。



(3) 前処理の設定・実行

分析に先立っては、「形態素解析」と呼ばれる前処理が必要である。前処理には、形態素解析のための辞書が必要であり、KH Coder では、「MeCab」か「ChaSen」の 2 種類が用意されているが、ここでは「MeCab」を選択し、「OK」ボタンを押す。次に、[前処理 (R) - 前処理の実行] から前処理を実行する。前処理が完了すると、「Database Stats」内に前処理の結果が表示される。

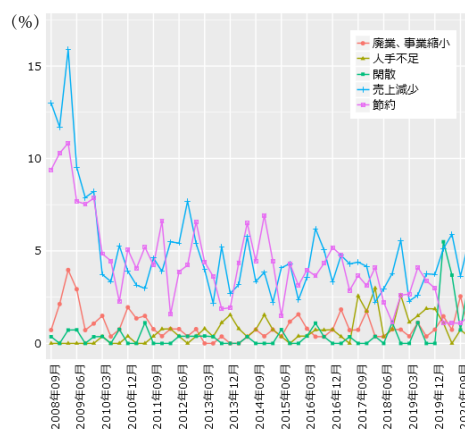


(4) 集計と結果の確認

[ツール(T) - コーディング - クロス集計] で、「参照」ボタンからコーディングルール・ファイルを選択する。ここでは、当課公式サイトで公表しているものを用いる。次に、クロス集計から「調査年月」を選び、「集計」を押す。集計が完了すると、各キーワードについて、調査月ごとのコメントの数と割合が表に出力される。

Year	*廃業、事業縮小	*人手不足	*閑散	*売上減少	*節約	ケース数
2008年09月	2 (0.72%)	0 (0.00%)	1 (0.36%)	36 (13.00%)	26 (9.39%)	277
2008年12月	6 (2.13%)	0 (0.00%)	0 (0.00%)	33 (11.70%)	29 (10.28%)	282
2009年03月	11 (3.97%)	0 (0.00%)	2 (0.72%)	44 (15.88%)	30 (10.83%)	277
2009年06月	8 (2.93%)	0 (0.00%)	2 (0.73%)	26 (9.52%)	21 (7.69%)	273
2009年09月	2 (0.72%)	0 (0.00%)	0 (0.00%)	22 (7.89%)	21 (7.53%)	279
2009年12月	3 (1.07%)	0 (0.00%)	1 (0.36%)	23 (8.21%)	22 (7.86%)	280
2010年03月	4 (1.49%)	1 (0.37%)	1 (0.37%)	10 (3.73%)	13 (4.85%)	268
2010年06月	1 (0.37%)	0 (0.00%)	0 (0.00%)	9 (3.35%)	12 (4.46%)	269
2010年09月	2 (0.75%)	0 (0.00%)	2 (0.75%)	14 (5.28%)	6 (2.26%)	265
2010年12月	5 (1.95%)	1 (0.39%)	0 (0.00%)	10 (3.91%)	13 (5.08%)	256
2011年03月	3 (1.35%)	0 (0.00%)	0 (0.00%)	7 (3.15%)	9 (4.05%)	222
2011年06月	4 (1.49%)	0 (0.00%)	3 (1.12%)	8 (2.97%)	14 (5.20%)	269
2011年09月	2 (0.77%)	1 (0.38%)	0 (0.00%)	12 (4.62%)	11 (4.23%)	260

また、右下の「すべて」ボタンを押すと、折れ線グラフでも結果を確認できる。



先程読み込んだコーディングルール・ファイルには、キーワードが5つ設定されているため、折れ線グラフにするとグラフの本数が多くなり一部の箇所が見づらい。

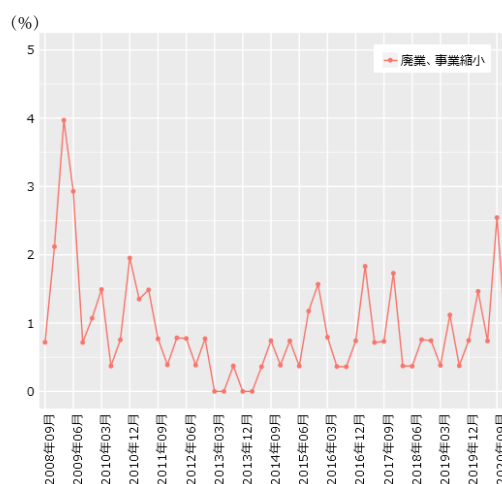
このような場合には、グラフが表示されているウィンドウの左下の「調整」ボタンから各グラフの表示・非表示を選択できる。また、集計表のウィンドウで「バブル」ボタンを押すと、キーワード数が多くなっても一度に全てのキーワードを比較して見ることができる。【図2】はバブルプロットといい、正方形の大きさに各調査月のキーワードの割合を表している。

5 分析結果について

各キーワードの折れ線グラフ及び【図2】から、各キーワードを観察する。

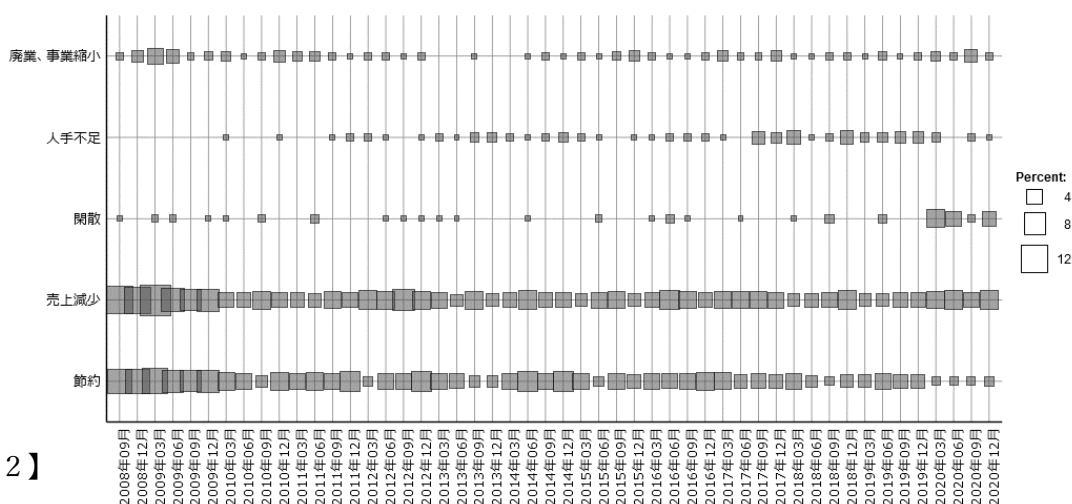
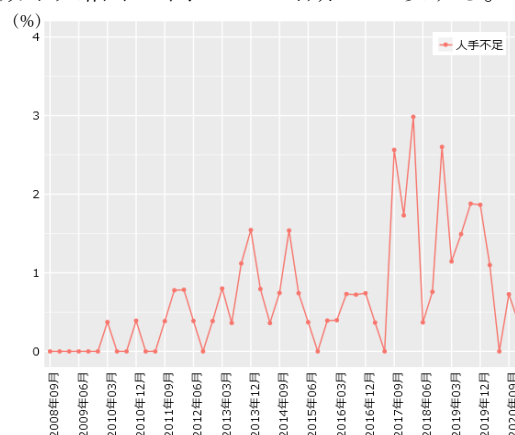
(1) キーワード「廃業、事業縮小」

リーマン・ショック後、DI値が過去最低となった平成21(2009)年3月前後の割合が高い。一方、新型コロナウイルス拡大期では令和2(2020)年9月は比較的高かったものの、その前後はほぼ平年並みである。



(2) キーワード「人手不足」

平成29(2017)年後半から令和元(2019)年までの期間で割合が高い。なお、これは有効求人倍率が高かった時期⁶に一致する。

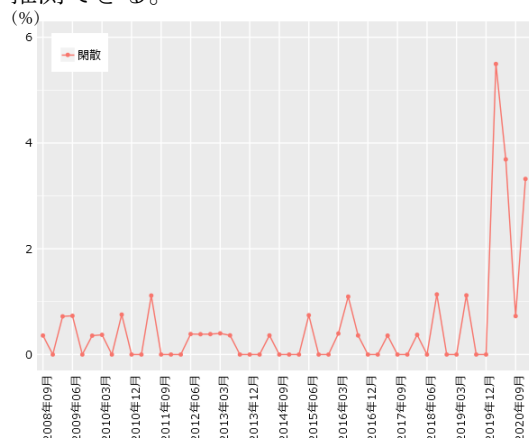


【図2】

⁶ 茨城県の有効求人倍率(季節調整値)は平成29年の後半から令和元年末頃までの期間は約1.6倍となっている。
https://jsite.mhlw.go.jp/ibaraki-roudoukyoku/roudoukyoku/gyoumu_naiyou/antei/jyousei.html (茨城県労働局)

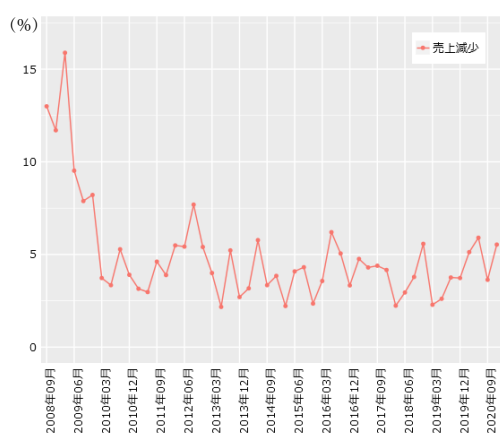
(3) キーワード「閑散」

令和元(2019)年まではあまり見られなかったが、令和2(2020)年の新型コロナウイルス拡大期に入り、割合が大きく上昇した。新型コロナウイルスが国内で流行し始めた3月や緊急事態宣言解除直後の6月はまだ人出は少なく、9月頃になって人出が戻ってきたと考えられる。しかし、その後再び感染が拡大し、外出自粛要請等も相まって人出が減り、「閑散」の割合が再び高まったと推測できる。



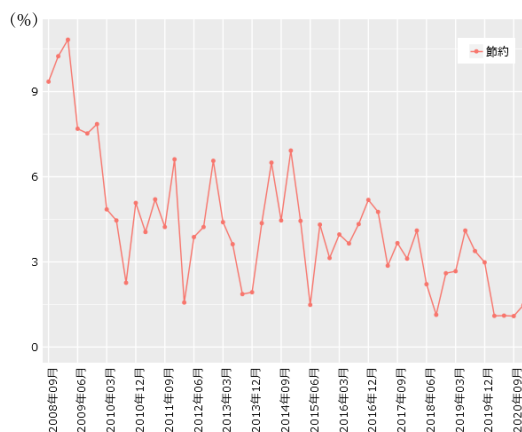
(4) キーワード「売上減少」

リーマン・ショック期で割合が高く、本調査のDI値が過去最低となった平成21(2009)年3月には最も高くなった。それ以外の期間では、概ね5%前後で推移している。



(5) キーワード「節約」

(4)の「売上減少」と同様に、リーマン・ショック期で割合が高く、平成21(2009)年3月には最も高くなった。一方で、新型コロナウイルス拡大期では低い割合となっている。



なお、「売上減少」と「節約」は似た動きをしているが、新型コロナウイルス拡大期では、「売上減少」に対して「節約」は低位となっており、傾向が異なっている⁷。



6 まとめ

ここまで、キーワードの出現割合を時系列で見ることで、あるキーワードがどの時

⁷ 「売上減少」と「節約」の令和2(2020)年3月以降の新型コロナウイルス拡大期を除いた期間での相関係数は0.73となり、強い相関が見られる。なお、新型コロナウイルス拡大期も含めた場合の相関係数は0.66となった。

期に多く見られたかを観察した。

リーマン・ショック期と新型コロナウイルス拡大期では、出現割合の高いキーワードに違いが見られ、リーマン・ショック期には「節約」や「廃業、事業縮小」に関するコメントが多かった一方、新型コロナウイルス拡大期では、「閑散」に関するコメントが多く、それぞれの時期の特徴を掴むことができた。

以上のように、テキストデータを時系列で俯瞰するには、今回紹介した方法が有効である。とりわけ、本調査はテキストの量が多く、全て読むには時間を要するため、どの時期にどのような特徴が見られるかを大まかに把握したい場合には特に有効となる。