

経済新聞記事からの予測語による自動車需要台数変動予測

小西 慶和† 中島 陽子‡ 本間 宏利‡

† 釧路高専 電子情報システム工学専攻

‡ 釧路高専 情報工学科

s130707@cc.kushiro-ct.ac.jp

1 はじめに

我々の生活において、景気の変動は様々な影響をもたらす。たとえば、一般的に景気が良い時は給料が増えたり、モノが売れやすくなるといった現象がおこる。反対に景気が悪い時は、失業者や会社の倒産が増加する。特に、企業や投資家にとって、景気の変動は多大な影響をもたらす。従って、商品の売れ行きや株価の変動を予測するために、景気動向を知ることは非常に重要である。

景気動向を予測する方法として、内閣府が発表している景気動向指数がある [1]。これは、生産、雇用など様々な経済活動での重要かつ景気に敏感に反応する指標の動きを統合することによって、景気の現状把握及び将来予測に資するために作成された指標である。しかし、景気動向指数から、現状の景気動向が知ることはできても、これを用いて将来の景気動向を予測することは極めて難しい。また、一般的にいうところの景気動向の予測とは、株価の変動や各人の経験則から予測している。機械的に景気動向を予測する、といった実用的なシステムは現段階では存在していない。景気動向予測に関する研究は行われてはいるが、予測するための根拠となる表現に対しての研究であるため、景気動向を予測するためには不十分である [3]。

そこで、本研究では、以上のことから景気動向を予測するシステム構築を目標にして、新聞記事から景気動向を予測する手法を提案する。

2 研究手法

本研究は、新聞記事から経済の景気動向を予測することを目的としているが、全ての経済に関わる指標を対象とするのは、範囲が広すぎて現実的ではない。例えば、建設業は東京オリンピック開催決定により活動が活発化していても、農業は悪天候により生産量が大幅に減少した、というような状況もある。このように、産業全域を対象にすることは望ましくない。従って、本研究では、日本の代表的な産業の1つである、自動車産業を対象を絞り、それに対する景気動向の予測を行う。自動車産業に関する景気では、自動車の販売台数や株価の変動など、様々な要因が考えられる。自動車産業の場合、国外への進出を行っているため、企業の利益や株価などの変動と比較しては、国内のみの正しい結果を得ることはできない。そこで、本研究では国内の自動車需要台数のみを対象とする。

過去の自動車需要台数のグラフから、自動車産業の景気変動の様子は容易に見つけることができる。好景気・不景気な時期を迎える以前の新聞記事中には、それらが到来することを示唆した内容の文が多く含まれていると予想される。そのような文の中には、他の時期には見られないキーワードとなるような単語が含まれていると考えられる。キーワードとなる単語を抽出し、他の時点でもその単語が含まれていれば、その先に大きな変動が来ると予想することができる。このように、他の時点よりも出現頻度の多い語を新聞記事から抽出し、それを基にして将来の予測を行っていく。

3 自動車経済記事の抽出

新聞記事 (毎日新聞 1994年～2000年) から自動車産業の景気変動を予測するにあたって、自動車経済記事を対象とする。そのため、あらかじめ全記事中から、経済タグによって経済記事のみを抽出し、さらに、それらに対し“車”というキーワードが含まれた記事を抽出することで、自動車経済記事を獲得する。しかし、“車”というキーワードで抽出した場合、“電車”や“乳母車”といった自動車に関係のないものも含まれてしまうため、明らかに関係のないものは、この段階で除去する。ここで抽出した自動車経済記事を基に、処理を進める。

4 予測語を用いた景気変動予測

本手法は、過去の自動車産業界の好景気・不景気時点の新聞記事より、出現頻度の多い語を抽出し、それらをもとに景気変動予測を行う。他の時点より出現頻度の多い語を“予測語”と定義する。予測語を用いた自動車需要台数の予測方法を述べる。

4.1 予測語の抽出

予測語は、過去の自動車産業界の好景気・不景気時の新聞記事から抽出を行う。従って、あらかじめ抽出しておいた自動車経済記事を好景気と不景気に分類する必要がある。新聞記事の分類は、過去の自動車需要台数の変動の様子を基に行う。

自動車産業の景気を表す要因は複数存在すると考えられるが、今回は日本国内の景気変動を予測するため、図1の国内自動車需要台数のグラフを参考にする[2]。図1から、自動車需要台数は1996年と1999年にピークを迎えていることがわかる。このピークの直前の新聞記事中には、それらの変動を示唆するような用語が数多く含まれていると予想される。従って、グラフが上昇中の1994年を、将来的に好景気になることから**P-予測記事集合 (P-記事)**、グラフが下降中の1997年を、将来的に不景気になることから**N-予測記事集合 (N-記事)**とする。

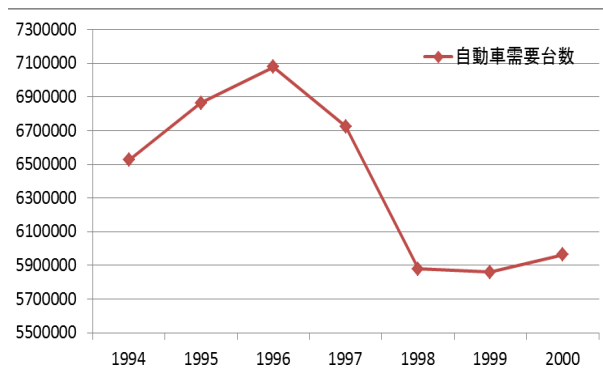


図1: 国内自動車需要台数の推移

次に、P-記事、N-記事から予測語を抽出する。記事中の用語は、どれも景気変動を示唆する言葉の可能性はあるが、特にサ変活用のある名詞が大きな関係があると考えられる。例えば、「国内生産は、政府の新政策により前年に比べ回復した。」という文が記事中に存在した場合、「回復した」という表現が一番景気変動に影響があると考えられる。他にも、「上昇した」や

「下降した」など、サ変活用のある名詞は物事の結果を表した表現であるため、予測語にふさわしい。従って、サ変活用のある名詞のみを予測語とし、これらを形態素解析を利用して抽出する。

4.2 予測語の分類

P-記事、N-記事からそれぞれ予測語を抽出し、それらを好景気・不景気を示唆する用語となるように分類する。ここで、好景気を示唆する予測語を**P-予測語**、不景気を示唆する予測語を**N-予測語**とする。分類は、P-記事とN-記事から獲得した予測語の出現頻度の差を算出して行うが、それだけでは不十分な場合が存在する。例えば、出現頻度の差はないが、意味的に明らかに好景気・不景気を示唆している用語は存在する。出現頻度の差だけで予測語を抽出した場合、それらは除外されてしまい予測の精度を落とすことになる。そこで、以下の3種類に対して分類を行う。

1. 出現頻度の差が多いもの
2. 出現頻度の差はないが、意味的に分類可能なもの
3. 出現頻度自体は少ないが、意味的に分類可能なもの

1.の方法では、P-記事とN-記事から抽出した予測語に対し、出現頻度の差が10という閾値を設定し、それ以上ならば正しい予測語として認識を行った。もともと、P-記事に含まれていたものはP-予測語、N-記事に含まれていたものはN-予測語とする。

2.の方法は、それぞれ記事中の出現頻度の高い用語であるため出現頻度の差は少ないが、意味的には間違いなく好景気・不景気を示唆している使われ方をしていものに対して行う。これらは、1.の閾値を満たさなかったものに対して目視により調査する。好景気を示唆している用語であればP-予測語、不景気を示唆していればN-予測語とした。

3.の方法は、これまでの方法とは異なり、元々出現頻度が低いものから予測語の抽出を行う。上記2つの方法は、元々出現頻度が高いものに対して用語を選別し抽出を行った。しかし、出現頻度は低くても明らかに好景気・不景気を示唆していること用語も記事中には含まれている。それらを抽出するために、目視により選別し抽出を行った。

表1は上記の方法で抽出された用語の例である。表1から、抽出された予測語は1のパターンが多いことがわかる。これは、出現頻度の差から獲得したものであり、単語の意味的に正しく分類されているかは考慮さ

表 1: 予測語の内訳

パターン	P-予測語 (個)	N-予測語 (個)
1	18	22
2	5	3
3	5	3
合計	28	28

れていない。P-予測語には、「協議」や「連続」、N-予測語には、「発表」や「消費」という単語が分類されている。一見すると、これらの単語は好景気、不景気を示唆している単語には見えないが、それぞれ大きな変動が来る前の時期に多く出現されていたため、予測に関係がある単語と考えられる。2のパターンからは、P-予測語に「増収」や「改善」、N-予測語に「操業」や「減少」が得られた。また、3のパターンからは、P-予測語は「増加」や「上昇」、N-予測語は「悪化」や「下落」が得られた。2と3のパターンは、どちらも得られた予測語の個数は少ないが、将来の変動を示唆しているような文で使われているため、より正確な予測を行うためには不可欠な単語である。

4.3 極性の付与

これまでは予測語の抽出を行ったが、抽出した予測語(サ変名詞)のみで景気変動を予測するのは不十分な場合がある。それらを解決するために文全体が好景気を示唆しているのか、不景気を示唆しているのかを表す**極性**を付与する。予測語は、それ単体で好景気や不景気を示唆しているが、共起する表現によっては極性が反転する可能性がある。例えば、「赤字は増加しなかった」とあった場合、「増加」という単語はP-予測語のため好景気を示唆しているが、「しなかった」という表現により、全く逆の意味を持つことになる。しかし、「赤字は」という表現が「増加しなかった」に係っているため、文全体としては好景気を示唆していることになる。このように、予測語の後ろにくる動詞部分と、予測語に係っている名詞にも着目する。この時、予測語の直後に位置する動詞を**極性動詞**、予測語に係っている名詞を**極性名詞**と定義する。予測語は、あらかじめP-予測語とN-予測語に分類していたため、それらを重みとして用いる。極性動詞は、「～した」や「～となった」などの肯定的なものにはP、反対に「～しなかった」や「～ない」などの否定的なものにはNの重みを付加する。極性名詞は目視により、「プラス」や「増」などのポジティブなものはP、「マイナス」や「減」な

どのネガティブなものはNとした。これら3つの指標から、文全体の極性を付与する。次の8パターンにより、極性付与を行う。

1. P-極性名詞 × P-予測語 × P-極性名詞 ⇒ P-文
2. P-極性名詞 × P-予測語 × N-極性名詞 ⇒ N-文
3. P-極性名詞 × N-予測語 × P-極性名詞 ⇒ N-文
4. P-極性名詞 × N-予測語 × N-極性名詞 ⇒ P-文
5. N-極性名詞 × P-予測語 × P-極性名詞 ⇒ N-文
6. N-極性名詞 × P-予測語 × N-極性名詞 ⇒ P-文
7. N-極性名詞 × N-予測語 × P-極性名詞 ⇒ P-文
8. N-極性名詞 × N-予測語 × N-極性名詞 ⇒ N-文

前述の「赤字は増加しなかった」の例は6.のパターンにあたるため、文全体はP-文を意味する。

4.4 景気変動の予測

本手法では、自動車経済記事を対象に、文の極性を付与した文の出現頻度数から景気動向の予測を行う。極性を付与した文の出現頻度数と、国内自動車需要台数に強い正の相関が存在すると予想される。その根拠として、あらかじめ抽出しておいた自動車経済記事に対して、記事全てに極性の付与を行った結果と、過去の国内自動車需要台数の結果との一致が挙げられる。まず、あらかじめ用意していた自動車経済記事を1年単位に分割し、それらに対して記事中の文の極性を調べた。記事全体のP-文の総数とN-文の総数の差をとったグラフを図2に示す。

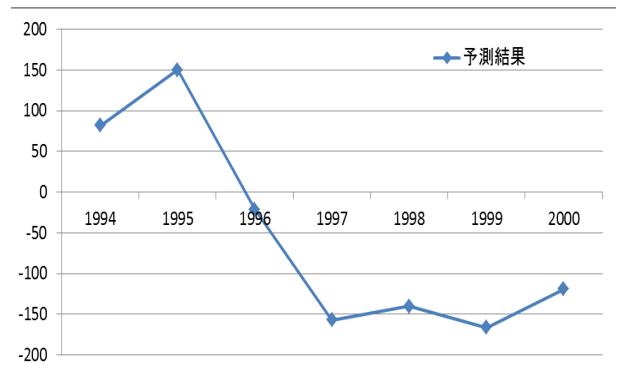


図 2: 極性の合計

図1と図2のグラフはそれぞれy軸の変数が異なる。今回はグラフの形状の一致から、本手法と自動車需要台数の関係性を見つけるため、それらを比較する

ためにグラフを統一する。そこで、極性を付与した結果のグラフを、自動車需要台数のグラフに合わせることにする。今回は、自動車需要台数の最大値と最小値とのふり幅が一致するように、値を補正する。

$$\text{倍率} = \frac{\text{最大値} - \text{最小値 (需要台数)}}{\text{最大値} - \text{最小値 (抽出結果)}} \quad (1)$$

$$\text{底上げ値} = \frac{\text{平均値 (需要台数)}}{\text{倍率}} - \text{平均値 (抽出結果)} \quad (2)$$

本手法で抽出した結果は、負の値が含まれているため、単純に倍率だけ合わせることは不可能である。従って、(2)式により負の値がなくなるよう底上げを行う。それに対して、(1)式の倍率の変更を行うことでグラフの統合を図った。その結果が図3のグラフである。

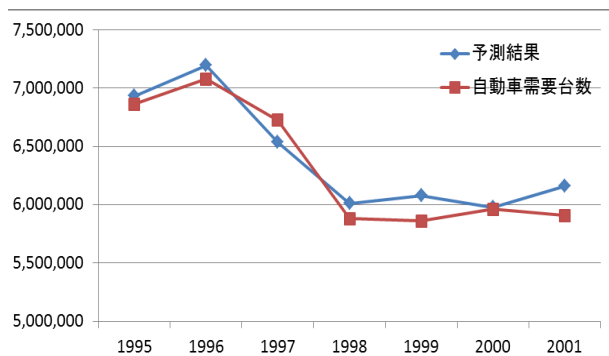


図 3: 自動車需要台数との比較

景気予測結果であるが、本研究では1年後の予測を想定しており、各年で抽出された結果は、それぞれ翌年の自動車需要台数と比較することで結果を確認する。従って、1994年から2000年までの記事から予測を行ったため、結果は1995年から2001年となっている。図3の結果から、多少の誤差はあるが、グラフの大まかな形状は一致していると考えられる。そこで、本手法は自動車需要台数を予測できているとし、実際に毎日新聞記事(2012年)1年分のコーパスに対して予測を行った。図4が予測結果であり、予測した値と2013年の自動車需要台数の値はほぼ一致していることがわかる。以上のことから、本手法では自動車需要台数の予測に成功したという結果を得ることができた。

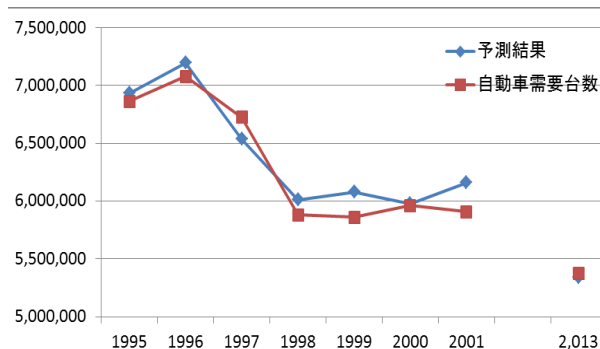


図 4: 2013年予測結果

5 おわりに

本研究は、過去のデータとの比較から、グラフの相違により変動の予測を行った。今回は、国内自動車需要台数のデータとの比較を行い、グラフの大まかな形状の一致から自動車需要台数の予測ができたといえる。しかし、抽出した予測語から必ずしも正しい結果が得られるわけではなく、予測語の抽出方法や極性の付与方法などによって、結果が大きく左右する可能性がある。予測語のみでは不十分な場合が存在するため、今回は極性の付与を行ったが、極性付けされたP-文は0.86%、N-文は0.75%と極性が付与されていた。しかし、予測語の抽出が自動化されていなかったり、極性の付与を行っていない文も存在するという問題がある。これらは、今後新たな規則の適用によって解決し、より精度を上げていく必要がある。

今回は、自動車需要台数のみに着目したが、今後は自動車産業に関する他のデータに対しても研究を進めていきたい。そして、将来的には自動車産業のみではなく、様々な産業のデータに対しても予測をすることで、結果的に日本全体の景気変動を予測したいと考えている。

参考文献

- [1] 内閣府 HP, “<http://www.esri.cao.go.jp>”
- [2] 日本自動車工業会, “<http://www.jama.or.jp/>”
- [3] 谷口翔太, 坂地泰紀, 酒井浩之, 増山繁. 経済新聞記事から抽出した景気動向を示す根拠表現への極性付与手法の提案. 言語処理学会第16回年次大会発表論文集, 2009.