単語極性情報を用いた未来動向予測手法の提案

中島 陽子[1]　高木景矢[3]　ミハウ プタシンスキ[2]　本間 宏利[1]　桝井 文人[2]

A Proposal of Prediction Method Using Word Polarity Information for Future Event Prediction Support System

Yoko NAKAJIMA, Keiya Takagi Michal PTASZYNSKI, Hirotoshi HONMA, Fumito MASUI

Abstract: In recent years, there has been an increased demand for future prediction in relation to social affairs, progress in science and technology, and economic circumstances. Moreover, there are large amounts of text data on the web, and there has been research into methods of future prediction support using natural language processing technology aimed at this. In previous research, we confirmed the effectiveness of future prediction supporting sentences, using an answer model generated through the use of pattern combination-based machine learning that considers language processing in word order for sentences referring to future events (FRS) using a newspaper corpus as learning data. However, there is other effective information on the Web in addition to just newspaper corpus, and this can also be used for sentence supporting prediction. Additionally, a method of improving prediction accuracy is to prepare an FRS classifier for each domain considering the sentence characteristics of each news domain, and acquire sentences supporting prediction. In this research, we obtained prediction support sentences related to future events from the news corpus on the Web, proposed a future event prediction support method using word polarity information, and showed that the prediction results exceeded the results of previous experiment.

Key words: 自然言語処理，将来言及文，極性情報，未来予測

1　はじめに

現在，Web を媒介として，大量のテキストドキュメント (新聞，ブログ，ツイッター等) が容易に入手可能であり，これらのデータを対象に自然言語処理技術 (情報抽出，テキストマイニング等) を利用した未来予測手法が研究されている．これらの研究には，株式市場の予測 [1, 2, 3] や選挙当選予測 [4]，医療分野 [5] に焦点を当てた研究は行われている一方，汎用的なイベント予測に関する研究は少ない．

Nakajima at el. [6, 7] は，ニュースコーパスを用いて未来イベントに言及している文 (将来言及文) を学習データとし予測モデルを生成する手法で将来予測支援システムを目指しており，未来インベント予測において将来言及文の有効性を明らかにしている [8]．予測精度を向上させるために，ニュース分野ごとに，文の特徴が異なることから，分野別の将来言及文分類機を生成し実験することを提案している．我々は，分野を科学技術，政治経済，国際の 3 分野に拡張し，汎用的に利用できる予測エンジンを目指す．Nakajima et al. [9, 10] の予測手法は，将来言及文が未来予測に有用かを確認するためのプロトタイプ実験を行なっている．予測したいイベントの関連文をパターンベースの機械学習モデル SPEC [11] で学習させ解答モデルを生成する手法を採用していた．人々が未来イベントが起きるか起きないかについてニュースや専門家の意見を参考にして予測する場合 “そのイベントが起きる"という問題を考えるとき，参照する記事集合の極性 (ポジティブまたはネガティブ) は考えるための情報のひとつである．評判分析や株価予測などの分野では，自然言語処理ベースの極性を用いた予測方法が研究されており，高い精度を得られるようになってきている [3, 12]．しかし，極性情報を用いた手法が未来イベントの予測に応用されている例はまだ少ない．

そこで，我々は汎用的な未来予測支援システムを構築するために，Web から科学，経済，国際の各分野ごとに，有用な予測支援文の獲得と単語極性情報を用いた未来動向を予測する手法を提案する．なお本論文は，nakajima et al. [13] で発表した内容に考察を拡張したものである．

次の章では将来言及文の定義について述べ，3 章では，単語極性値を用いた未来イベント予測手法を提案する． 4 章では予測実験， 5 章では実験結果について報告し， 6 章で実験結果について議論し，最終章でまとめを述べる．

2　将来言及文

本研究において，将来へ言及している文を未来動向の際の予測支援文として予測に用いる．将来言及文はその文がその時点よりも未来へ言及している文と定義する．将来へ言及する文にはいくつかパターンがあるので，その例を示す．

1. 三菱化学では 2025 年前後に Li イオン 2 次電池を搭

[1]釧路工業高等専門学校 創造工学科
[2]北見工業大学 情報システム工学科
[3]豊橋技術科学大学 情報・知能工学系

載した HEV の本格普及が始まるとみている。

2. 各国から派遣された総勢約 500 人は 23 日に担当地域に移動して，対立政党間で脅迫などがないか監視する。

3. 年末の税制改正大綱の決定に向けて，今後，調整が本格化することになります。

4. 長期的には半導体事業のうちパソコン用などの汎用 DRAM 事業 の割合を減らしていく方針。

5. 国内では１２工場を７工場に集約し，８００人規模の希望退職を募るなど大規模なリストラに踏み切る。

6. アメリカの部品やソフトウエアの輸出，技術の移転を制限すると発表しました。

1)，2)，3) のように未来の時間情報（年，月，日，来年，来月，未来）を明記しているパターン，4) のように「目的」など未来を明確に言及する言葉が含まれている文章，5) のように未来の時間情報や未来を言及する言葉が含まれていないが未来の動向に言及している．6) では，主節は過去形であるが，内容は未来の動向に言及している文である．

新聞記事や専門家の意見，政府の計画などは，意図的な「フェイクニュース」でない限り，専門家やその情報に詳しい記者，あるいは有識者が選択的に書いたものである．正当なニュース記事については，いくつかのチェック機構が設けられている．

これを考慮すると，未来に言及した文章は社会情勢や過去の事実の背景，専門家の調査結果や意見，研究動向，時事的な事実など素人では調査が困難な知識背景に基づいて表現されていると考えられる．

nakajima et al. [10] によって，将来言及文を用いた未来動向予測は，1 年間に収集した全ての情報を元に人間が予測するよりも，16〜30 文程度の将来言及文を提供すれば高い予測精度が得られることが明らかにされている．彼らの実験では，将来言及文の数については予測結果に差がないことが示されている．

さらに，ニュースコーパスから分野を指定せずに将来言及文を収集するよりも，そのイベントの分野に適した将来言及文を収集することが重要だと言及している．

本研究では，予測支援文の収集範囲を Web 上に存在するデータに拡大し，さらに，科学技術，政治経済，国際の各分野に対応する将来言及文分類モデルを生成し，予測支援文を獲得する．分野別に自動分類するモデルにより予測支援文が得られれば，汎用的かつ予測精度を向上させることができると考える．将来言及文分類モデル生成アルゴリズムは，nakajima et al. [7] が提案したものを採用する．

3 提案手法

本研究における将来のイベント予測は，予測したいイベントの質問文と可能性のある複数選択肢をテキスト入力し，そのイベントに関連する文に対して単語極性情報に基づいてスコアを算出して解答を選択する．関連記事を収集するためのキーフレーズの生成，単語の極性情報を利用した得点計算の選択肢とそれを利用した出題予測方法について述べる．

将来のイベント予測の入力例は以下に示す．

予測イベント：
2010 年に行われる参議院選挙の結果を予測せよ．

選択肢：

1. 民主党が単独過半数をとる．

2. 民主党だけでは単独過半数を取れないが，民主党を含む与党で過半数をとる．

3. 民主党を含む与党で過半数を確保できない．

提案手法の以下に示す手順で実現する．

1. 予測したいイベントを問題文と選択肢として与え，各文の名詞・動詞を抽出しキーフレーズ集合とする．
2. キーフレーズ集合を用いて Web やテキストコーパスからイベントの関連文を取得する．
3. 関連文を将来言及文分類モデルに入力し，将来言及文し予測支援文を生成する．
4. 予測支援文に含まれるキーフレーズ頻度と文の単語極性値を用いて解答選択肢のスコアを算出する．
5. 予測結果はスコアが最も高い選択肢を解答とする．

3.1 予測支援文の取得

予測支援文は具体的には，予測したいイベントに関連する Web やコーパスから関連する文であり，未来動向に言及している文の集合とする．

予測支援文の取得手順は次に示す．

1. 問題文と解答選択肢を形態素解析し，名詞と動詞および時間を表す単語を抽出しキーフレーズ集合を生成する．
2. Web を介し，キーフレーズ集合の要素が同時に 2 つ以上を含む記事を予測イベントの関連記事として取得する．
3. 獲得した関連文を将来言及文分類モデルに入力し，その結果を予測支援文とする．

ここで，使用する将来言及文分類モデルの性能は，政治経済学分野で 0.91，国際分野で 0.95，科学技術分野で 0.98 である．また，形態素解析は，形態素解析器 MeCab [1] で行う．

3.2 単語極性値

単語における極性とは，ある文章が良い印象（ポジティブ）を与えるか，悪い印象（ネガティブ）を与えるかを表す情報である．本研究では，極性を表す数値を極性値と定義する．単語極性の判定は日本語評価極性辞書 [14, 15] を用いる．日本語評価極性辞書では，頻度が高い単語を集めた約 5000 の評価表現と kawahara et al. [16] の 5 億文コーパスから頻度の高い名詞を抽出し評価極性を割り当てている．日本語評価極性辞書では例えば，ポジティブの極性の単語に「賛成」は登録されているが,「賛成」の対義語である「反対」がネガティブの極性の単語には登録されていないなど本実験を行う上で調整が必要である．そこで，本実験では手動でポジティブ極性に 11 個の

[1] http://taku910.github.io/mecab/

単語を追加し，負極性に 17 個の単語を追加した．追加した単語は，日本語評価極性辞典に登録さている単語の反意語と同義語である．今回追加した単語を以下に示す．

ポジティブ極性

完成 懸命 高い 勝利 目標 優勢 必ず 実績 支持 目指す 搭載

負極性

反対 抗議 否定 不成立 劣勢 失望 反発 解散 悩む 低迷 難航 下回る 不満 批判 再燃 対策 悠長

3.3 選択肢のスコア計算

次に，予測したい未来イベントの予測問題文，選択肢，予測支援文と極性情報を用いた予測手法について述べる．解答を選択するための選択肢のスコアは予測支援文，問題文および選択肢を用いて次に示す 3 値を用いて算出する．

- 極性値
- 分類モデルの出力で得られる将来言及の強さを示す値
- 予測支援文中のキーフレーズ出現頻度

選択肢の極性値を P_c，予測支援文の極性値を P_s，将来言及の強さを示す値を C，予測支援文と選択肢に共通で含まれるキーフレーズの個数を n とする．予測支援文 1 文にあたえられるスコア S_i は，式 1 で表される．Ikeda at el. [17] は，極性シフタモデルが特徴関数とシフタモデルのパラメータの積で表されることを実証している．i は予測支援文の番号であり，予測支援文の個数を I とすると $1 \leq i \leq I$ となる．さらに，c は，文章が未来を暗示しているかどうかの指標であり，将来言及文分類モデルを用いて分類する際に，将来言及文で使用される形態素パターンの使用頻度と，文章中で使用される形態素パターンの構成要素に基づいて算出される． キーフレーズの個数 n は予測したい内容との関連度とする．

$$S_i = P_c \times P_s \times C \times n \tag{1}$$

予測支援文の数を I とすると選択肢のスコア S_x は 2 式で表される．x は選択肢番号であり，選択肢の個数を X とすると $[1 \leq x \leq X]$ である．

$$S_x = \sum_{i=1}^{I} S_i \tag{2}$$

予測支援文 1 文のスコア S_i の計算手法を詳細に述べる．

Step1: 選択肢の極性 P_c を日本語評価極性辞書を参照しポジティブの場合は 1，ネガティブの場合は -1 とする．また，構文が主節と従属節で構成されている場合，接続詞，接続助詞 (“また", “従って"など) がある場合，それらを境界とし文節に分割して適用する．分割数を A とすると分割されたそれぞれの文のスコアを S_{x_1}，S_{x_2}，$\cdots$ S_{x_A} で表す． Figure 1 の選択肢 2 のように接続助詞がある場合は,「民主党だけでは単独過半数をとれないが」と「民主党を含む与党で過半数をとる」の 2 文節に分けてそれぞれの文でスコアの計算を行う．接続詞や接続助詞 (“しかし", “一方"など) の極性を反転させる反転子については文節に分けることで考慮可能になる．

Step2: 次に予測支援文 1 文のスコア S_i を 1 式により算出する．Figure 1 は選択肢 1 を例にスコア S_1 を計算した例である．説明のため予測支援文の将来言及の強さ C は将来言及分類モデルで 2.0 が算出されたと仮定し例を示す．予測支援文 1 文のスコアは S_1 は $[S_1 = P_c \times P_s \times C \times n = 1 \times (-1) \times 2 \times 2 = -4]$ となる．

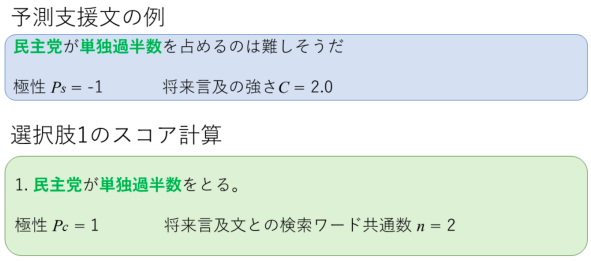

Figure 1: 予測支援文を用いたスコア S_1 の計算方法

Step3: 各選択肢のスコア S_x は全支援文 $[1 \leq i \leq I]$ の S_i を求めた後，総和を求める．ただし，分割した文節ごとに計算をした選択肢のスコアは各文節を統合し，スコア S_{x_1}，S_{x_2}，$\cdots$ S_{x_A} の和を S_x とする．Figure 2 に示すように 2 式より選択肢 2 のスコア S_2 は $[S_2 = S_{2\_1} + S_{2\_2} = -4 + 10 = 6]$ が算出される．

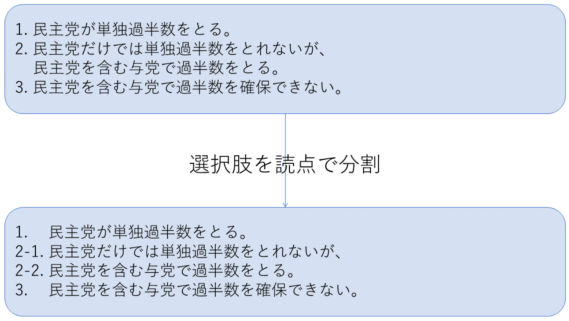

Figure 2: 文節を考慮した選択肢のスコア計算方法

Step4: 最もスコアが高い選択肢を予測結果とする．

以上の手順に従い，全ての予測支援文と解答選択肢に適用し未来動向予測問題の予測結果を得る．

4 実験

まず，ニュース分野別将来言及文分類モデルがそれぞれの分野における将来動向予測に有効であることを確認するために予備実験を行い，その後，本手法を用いた将来動向予測実験を行う．

4.1 実験設定

未来動向予測問題は，第 4 回先見力検定[2]の 2011 年–2012 年のイベントを予測する問題から抜粋した 7 問と今回の実験のために作成した科学技術分野に関連した 2015 年

[2]http://genseki.a.la9.jp/senken/index4.html

のイベントを予測するためのオリジナル未来予動向測問題 8 問を用いて実験を行う．

先見力検定は公益社団法人言語責任保証協会が実施しているもので，公共性の高い人（経営者，政治家）や市民生活に影響を与える意思決定を行う人を支援することを目的としてる．

問題の例を以下に示す．

第 4 回先見力検定抜粋問題の例

問題： 2010 年に行われるアメリカ中間選挙の結果はどうなるか．

選択肢：

1. 上院・下院とも民主党が過半数をとる．
2. 上院は民主党，下院は共和党が過半数をとる．
3. 上院は共和党，下院は民主党が過半数をとる．
4. 上院・下院とも共和党が過半数をとる．

オリジナル未来動向予測問題の例

問題： 2015 年には日本国内の「スマートフォン」の普及率は 50%を越えるか．

選択肢:

1. スマートフォンの普及率は 50%を超えるだろう．
2. スマートフォンの普及率は 50%を下回っているだろう．
3. スマートフォンの普及率は 50%を越え,70%に達しているだろう．

予備実験と提案手法実験で用いる未来動向予測問題は同じものとし， 問題文に対する解答選択肢は 2 つ以上の複数とする．

4.2 予備実験

予備実験はニュース分野の科学技術分野において，Nakajima et al. [8] のプロトタイプ手法を用いて予測実験により検証を行う．未来動向予測問題の関連記事を得るために第 4 回先見力検定抜粋問題には 2009 年の毎日新聞[3]コーパス，オリジナル未来動向予測問題の実験は 2012 年と 2013 年の毎日新聞コーパスを用る．

キーフレーズ

キーフレーズは形態素情報と意味役割情報を用い 3 単語組みとする. 問題文に形態素情報と意味役割情報の ラベルの付与を行い，そのラベルの組合せの条件は次に示す．

- 3 つのラベルのうち時間または数値のラベル 1 つと動詞，名詞，状態変化ありのラベル 1 つが含まれていること
- 1 の条件が満たさない場合は時間，数値，動詞，名詞および状態変化ありのラベル 1 つが含まれていること
- 問題文に付与されたラベルが 3 つの場合は 3 単語の全組み合わせ

将来言及文抽出

Web 上に存在するニュース記事から科学技術に関する将来言及文 500 文，そのほかの文 500 文を目視で収集し nakajina et al. [7, 8] の手法により分類モデルを生成する．

予測

各問題ごとに前述したコーパスからキーフレーズを用いて関連文を取得し，分類モデルにより予測支援文を取得する．将来言及文を入力データとして SPEC による学習を行うことで選択肢を選択するための解答モデルを生成する．選択肢は解答モデルに入力され，最大スコアを持つ選択肢が解答として選択される．

結果

予備実験の結果結果，第 4 回先見力検定抜粋問題の正答率は 42.8%，オリジナル予測問題の正答率は 62.5%であった．予備実験で使用した予測支援文は，問題の分野に限らず科学技術分野の将来言及文分類器を用いた．科学技術分野の問題を含むオリジナル予測問題の予測結果の方が正答率が高いことが示された．従って，科学技術分野の予測支援においてはイベントの分野に適した分類モデルを用いることが未来動向予測の精度向上に有効であることが示された．

ここで定義したキーフレーズでは予測支援をするための文数に差が出ること，少ないもので 10 文程度となってしまうことがわかった．提案手法ではキーフレーズ抽出の条件を緩和する．

4.3 提案手法による実験

分野別で生成する分類モデルはその分野の将来言及文の取得に有効であることを確認できた． 従って，政治・経済と国際分野のそれぞれの将来言及文分類モデルを生成し，単語極性情報を用いた未来動向予測実験を行う．本実験では各分野ごとの精度を確認するために予測する問題は国際，経済および科学技術分野の 3 分野について実験する．

未来動向予測問題の関連記事の取得には第 4 回先見力検定抜粋問題には 2009 年の佐賀新聞[4]と日本政府の経済白書〜[5]および毎日新聞コーパスを用いる．オリジナル予測問題には 2009–2010 年の佐賀新聞と経済白書，2009 年，2012 年および 2013 年の毎日新聞のコーパスを用いる．関連記事は Web 上から情報の抽出を行う python ライブラリとして 提供されている Beautiful Soup4[1] を用い，キーフレーズを google 検索機能に自動的に入力し Web を介して関連記事を取得する．取得した関連記事は各分野（国際，経済・政治，科学・技術）の分類モデルを用い，予測支援文として抽出する．

5 実験結果

各予測問題の予測実験の結果を Table 1 に示す．予測結果と実際の解答が一致した場合は **T** (True)，予測結果と実際の解答が一致しない場合は **F** (False) と表している． また，比較のために，予備実験の結果を示す．

単語極性情報を用いた未来動向予測実験の結果は第 4 回先見力検定選抜問題において 7 問中 5 問：正答率 71.4%，オリジナル予測問題において 8 問中 6 問で正答率 75.0%の精度で予測できた．

各予測問題の分野 (Table 2) ごとの正答率を示す (Table 3).

6 考察

第 4 回先見力検定抜粋問題において，従来のプロトタイプ手法による予測実験の結果が 47.5%であったのに対し，本提案手法では 71.4%となり，23.9 ポイントの改善が見られた．また，オリジナル予測問題においては 62.5%から 75.0%へと 12.5 ポイントの改善が見られた．

一見すると，精度の面ではそれほど良い結果ではないように思えるかもしれないが，すでに実施されている第 4 回先見力検定を受けた受験者の平均正答率は約 30%であった．この結果は解答期間 1 年間のうちにあらゆる方法でデータを調査し，受験した結果である．また，Kurokawa et al. [18] が数回実施した先見力検定の結果では，数回の平均正答率が 30%台であったことが言及されている．従って，提案方法において，正答率が 71.4%であったことは人間の予測能力の 2 倍以上の精度を示しており，未来動向を予測するのに有用な方法と考えられる．

[3] http://www.nichigai.co.jp/sales/mainichi/mainichi-data.html

[4] https://www.saga-s.co.jp/

[5] http://www.kantei.go.jp/jp//hakusyo/

Table 1: 単語極性情報を用いた手法による各問題の正誤の結果と，予備実験結果の比較

Question No.	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Accuracy in 提案手法の結果 [%]	Accuracy in 予備実験の結果 [%]
The 4th FPCT	F	T	T	F	T	T	T	-	71.4	47.5
OFPQ	T	T	T	T	F	T	F	T	75.0	62.5

Table 2: 各問題の分野: 科学技術 (ST), 政治経済 (PE), 国際 (I)

Question No.	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
The 4th FPCT	PE	I	I	PE	PE	PE	PE	-
OFPQ	ST	ST	ST	ST	ST	ST	ST	ST

Table 3: 分野別正答率: 科学技術 (ST), 政治経済 (PE), 国際 (I)

Field of question	ST	PE	I
Accuracy	0.75	0.60	1.00

正答率は先行研究よりも向上したが不正解となる問題の改善点を見つけるために不正解であった問題と選択肢および予測支援文から考察を行う. 問題文，選択肢および予測支援文の一部を示す．予測支援文抜粋における文中の下線はキーフレーズを示している．

予測問題文：
2010 年 6 月末時点で永住外国人への地方参政権付与が可決成立しているか.

選択肢：

1. 永住外国人への地方参政権付与が可決成立する．

2. 永住外国人への地方参政権付与が可決不成立である．

正解： 2

予測支援文抜粋に示す予測支援文の a, b, のように有用である文も取得されている一方で，d, e, g, j のように明らかに予測したい問題とは関係のない文が含まれている．キーフレーズの条件を問題文と選択肢に含まれる時間情報，名詞および動詞からなる 2 単語組みとしたことで関連性のない文が多く取得されてしまったために関連性のない文が多くなったと考えられる．

この問題は抽出した関連記事から TF-IDF(Term Frequency-Inverse Document Frequency) などで文章の特徴語を抽出する手法を適用し，時間表現と組み合わせることで貢献しない名詞や動詞が排除できると考える．

例えば，Yahoo!JAPAN[6]が提供しているキーフレーズ抽出 API では「2010 年 6 月末時点で永住外国人への地方参政権付与が可決成立しているか」のキーフレーズは，“地方参政権付与", “永住外国人", “可決", “2010 年 6 月末時点" および “可決不成立"をキーフレーズの重要度とともに取得できる．

支援文例 a, b, f では予測したい問題の関連文の中には不正解である選択肢と同様な内容の文章が見られた．関連記事を取得する際に用いるニュースサイト，専門サイトや専門家が執筆したサイト等に拡張することで異なる意見の記事も取得可能になるような改善が必要である．

[6]https://www.yahoo.co.jp/

予測支援文抜粋

a.永住外国人地方参政権付与案を今国会に議員立法で提出したい考えを伝えた。
b. 公明党は永住外国人に地方参政権を与える法案の国会提出を検討している。
c. 補正予算案は 13 日に衆院を通過し，遅くとも6 月12 日に成立する。
d. 買い取りは2010 年 6 月末まで続ける．
e. 捜査本部は2010 年7 月の時効成立をにらみ捜査員を増強して男らの情報を収集。
f. 小沢氏は永住外国人への地方参政権付与法案について早期実現に意欲を示した。
g. 首相問責決議の可決は昨年６月の福田康夫首相に対する決議に続いて２回目。
h. １９日の本会議で可決，成立する見通し。
i. ８日にも参院本会議で可決，成立する見通し．
j. タクシー適正化・活性化法が６月に成立し，１０月１日施行の見込み。
k. 議会多数を占める自民党が主導していることから会期中に可決成立する見通し。
l. 同日午後の衆院本会議で３分の２以上の賛成で再可決し，成立させる。

また，“成立する"という内容を暗示している c, e, g, h, i, j, k, l はキーフレーズで重要な “永住外国人"および “地方参政権付与"を含んでおらず，別の議案についての文章である．この問題は，キーフレーズに重要度を付与することで改善できると考える．

本実験に使用した問題のうち不正解の問題の共通点は選択肢に “6 月", “2015 年"のような数値の情報が含まれていた. 記事取得の際に取得対象とするのはキーフレーズの数値と同じ数値が含まれている場合のみであり，問題文が「80%に達するか」の際に「90%に達する」のような記事を取得することができない．従って本来未来予測に有用な予測支援文を取り逃している．数値や数値に付随する語を含めてキーフレーズに数値がある場合の対応を行う必要があると考える．

7 まとめ

本研究では，将来言及文分類モデルを 3 つニュース分野に拡張し，予測イベントの関連記事を検索するためのキーフレーズ定義し，文中の単語極性情報を用いて選択肢にスコア計算する手法により未来動向予測手法を提案した．本実験では，1–2 年後の未来動向を第 4 回先見力検定とオリジナル予測問題を用いて，それぞれ 71.45%, 75.0%の正答率で予測することができた．評判分析や株式市場分析に用いられている文の極性情報は未来イベント予測する場合にも有用性を示した．

提案した手法では日本語評価極性辞書の極性としたが，予測支援文を収集する際に日本語極性評価辞書にない単語を自動的に抽出し極性情報を付与するなどの改善， また， 未来動向予測に有用な予測支援文を取得するためにキーフレーズを洗練することで未来動向予測精度の向上が期待できると考える．

今後は nakajima et al. [10] が提案する形態パターンを用いた

機械学習による予測モデルと単語極性情報を併用するなど予測精度の向上を目指し，マーケティングやエネルギー需要予測など実世界に応用できる，また，多言語にも対応する未来予測支援システムの運用を目指す．

謝辞

本研究は JSPS 科研費 17K00324(2017–2019) の助成を受けたものである．

References

[1] K.Shirata, H. Tsuda, “The Prediction of Stock Market by Natural Language Processing and Deep Learning,” The Harris science review of Doshisha University, vol. 59-3, pp.163–172, 2018.

[2] H.Maekawa, T.Nakahara, K.Okada, et al., “Textual analysis of stock market prediction using breaking financial news : The azfintext system,” Operations research as a management science research, vol.58(5), pp.281–288, 2013

[3] S.Merello, A. Picasso Ratto, Y. Ma, et al., “Investigating Timing and Impact of News on the Stock Market,” IEEE International Conference on Data Mining Workshops (ICDMW), pp. 1348–1354, 2018

[4] P.Salunkhe, S.Deshmukh, “Twitter Based Election Prediction and Analysis,” International Research Journal of Engineering and Technology (IRJET), Vol.04-10, 2017

[5] H.A.SCHWARTZ, M.SAP, M.L.KERN, et al. , “PREDICTING INDIVIDUAL WELL-BEING THROUGH THE LANGUAGE OF SOCIAL MEDIA,” Biocomputing 2016, pp.516–527, 2016

[6] Y.Nakajima, M. Ptaszynski, H. Hirotoshi, F. Masui, “Application of future sentence reference extraction in support of future event prediction,” In: Proceedings of Workshop on Language Sense on Computer (IJCAI 2016), pp. 81-88, 2016

[7] Y.Nakajima, M. Ptaszynski, F. Masui, H. Hirotoshi, “A method for extraction of future reference sentences based on semantic role labeling,” IEICE Trans. Information and Systems, Vol.E99D, No.2, pp.514–524, 2016

[8] Y.Nakajima, M. Ptaszynski, F. Masui, H. Hirotoshi, “Future Reference Sentence Extraction in Support of Future Event Prediction,” International Journal of Computational Linguistics Research, Vol.9-1, pp.29–43, 2018

[9] Y.Nakajima, M. Ptaszynski, F. Masui, H. Hirotoshi, “Automatic extraction of future references from news using morphosemantic patterns with application to future trend prediction,” ACM Summer Newsletter. AI Matters, 2 (4) 13-15, 2016

[10] Y.Nakajima, M. Ptaszynski, F. Masui, H. Hirotoshi, “A Prototype Method for Future Event Prediction Based on Future Reference Sentence Extraction,” In: Proceedings of Workshop on Linguistic and Cognitive Approaches To Dialogue Agents (IJCAI 2017), pp.42–49, 2017

[11] M. Ptaszynsk, R. Rzepka R, K. Araki, Y.Momouchi, “SPEC-Sentence Pattern Extraction and Analysis Architecture,” In Proceedings of the Seventeenth Annual Meeting of the Association for Natural Language Processing, 2011

[12] M.Birjalia, A.Beni.Hssanea, M.Erritalib, “Machine Learning and Semantic Sentiment Analysis based Algorithms for Suicide Sentiment Prediction in Social Networks,” Procedia Computer Science, 113, pp.65–72, 2017

[13] Yoko Nakajima, Michal Ptaszynski, Hirotoshi Honma, Fumito Masui. A Proposal of Prediction Method Using Word Polarity Information for Future Event Prediction Support System. Advanced Informatics, Concept, Theory, and Applications (ICAICTA), No.95, 2019

[14] Nozomi Kobayashi, Kentaro Inui, Yuji Matsumoto, Kenji Tateishi. Collecting Evaluative Expressions for Opinion Extraction, Journal of Natural Language Processing 12(3), 203-222, 2005.

[15] M.Higashiyama, K.Inui, Y.Matsumoto, “Learning Sentiment of Nouns from Selectional Preferences of Verbs and Adjectives,” Proceedings of the 14th Annual Meeting of the Association for Natural Language Processing, pp.584–587, 2008

[16] D.Kawahara, S. Kurohashi. “A fully-lexicalized probabilistic model for Japanese syntactic and case structure analysis,” In Proceedings of the Human Language Technology Conference of the NAACL, Main Conference, pp.176-183, 2006

[17] D.Ikeda, H.Takamura, M.Okumura, “Learning to shift the polarity of words for sentiment classification (in Japanese), ” Artificial Intelligence Vol.25 No.1, pp.50–57, 2010

[18] T.Kurokawa, H.Kakeya, “Analysis of a tendency to answer in foresight official approval (in Japanese),” In:The Fifth media information inspection arts and sciences meeting, pp.50–55, 2009