将来言及文の分類精度向上を目的とした汎用型分類モデルの構築

中島 陽子[1]　本間 宏利[1]　Akmal Hakim[2]　ミハウ プタシンスキ[3]　桝井 文人[3]

Construction of general-purpose classification model for the purpose of improving classification accuracy of future reference sentences

Yoko NAKAJIMA, Hirotoshi HONMA, Akmal HAKIM, Michal PTASZYNSKI, Fumito MASUI

Abstract: In recent years, in order to suit to changes in social conditions with many uncertainties, research on future prediction and future strategies has been actively conducted in various fields. Most of them are research using statistical methods, and if an uncertainty element occurs, it is difficult to deal with linear statistical prediction. In previous research, classification models were generated using a method using morphological and semantic roll pattern in future reference sentences in news articles. However, news articles have domains such as economy and sports, and the method do not take into account that sentences patterns and the frequency of characteristic word differ for each domain. Therefore, in this study, we examine a method of determining sentence patterns and feature words for each news domain, propose sentence patterns that consider morphosemantic and feature words, and aim to improve the accuracy of classification models in each news domain.

Key words: 文章分類，自然言語処理，分類モデル，将来言及文，形態意味論

1　はじめに

社会情勢がめまぐるしく変化する今日，金融市場，軍事産業，経営経済などの多くの分野で精度の高い短期的および中長期的な未来動向予測の需要が高まってきている．高精度な未来動向予測の実現には，世界情勢の知識，歴史的経緯，有識者の見識など大量かつ多様な情報取得と専門的な分析手法が必要であり，研究者達の障壁となっていた．近年，インターネット網と SNS の普及により，様々な分野の膨大なテキスト情報が瞬時に容易に入手可能となったことで，自然言語処理技術を利用した未来予測研究が精力的に行われてきている．これらの研究には，株式市場の予測 [1, 2] や選挙当選予測 [3]，医療分野 [4] に焦点を当てた研究がある．しかしながら，これらの研究の多くは，膨大なデータと統計的手法を用いており，不確実性の高い突発的な出来事（イベント）には対応するのは難しい．さらに，予測対象分野を限定しており，問題固有の統計技法の他，専門的観点からのヒューリティクスを要することから汎用的観点からの研究は今後の課題とされている．

未来動向予測に必要な世界情勢の知識，歴史的経緯，有識者の見識など大量かつ多様な情報取得と専門的な分析手法を実現するために，中島ら [10] は，文を構成する意味役割情報を用いた形態パターンを利用し，ニュース記事などのテキストデータから獲得した将来イベントに言及する文を利用した未来動向予測支援システムの開発研究を行った．将来に言及する文は，未来動向予測の際に有用であることを確認したが，科学技術分野の未来予測に限定しており，また，将来に言及する文をニュース記事などから自動的に取得する精度向上の課題が残されている．

そこで，自動的に取得する将来言及文の精度を向上目的とし，意味役割情報の他に新たに未来動向を指し示す語を定義し，将来言及文を意味役割情報と未来語のパターンを機械学習により将来言及文の分類モデル生成と汎用性未来動向予測の実現を目指し，未来動向予測分野拡張を提案する．

2　先行研究

中島ら [5] は，将来に言及する文の意味役割情報と形態情報からなる文パターンが言語的にどのように表現されるかについて調査を行なった．彼らはこの将来に言及する文が一貫した言語的実体とみなすことができることを確認した．その後，Nie ら [6] はアラビア語と英語の将来言及文を分析し，中島らと同様の研究を行い，同様の結果を確認した．また，同様に Al-Hajj と Sabra[7] も単純な 1 単語パターンに限定して実験を行なっている．Yarrabelly ら [8] は，少なくとも英語で書かれたニュース記事に適用される場合，将来の参照の 1 つの予測文の抽出に依存関係を適用することも実行可能な方法であることを証明した．Hurriyetoglu ら [9] は，いくつかのヒューリスティックな時間情報を表す単語と単純な時間表現を適用し，Twitter でイベント発生までの時間を予測するために履歴コンテキストを解析した．イベントの展開のライブ追跡に適用できるが，この予備調査では，コンサートやサッカーの試合などのイベントのイベント発生までの時間の推定にのみ焦点を当てており，アプリケーションを評価するための優れた実証基盤であるにもかかわらず，そのような事前に計画されているイベント（サッカーの試合の日付は多くの場合固定されており，1 年先でもよく知られてる）の予測にとっては，それほど大きな課題ではない．た

[1]釧路工業高等専門学校 創造工学科
[2]釧路工業高等専門学校 情報工学科
[3]北見工業大学 情報システム工学科

だし，TTE 推定方法のみを考慮する場合，それらの研究は貴重な貢献であり，政治会議や結果が事前にわからない主要な市場変動など，より挑戦的なイベントの TTE の推定に適用可能である．

上述した研究においては，ニュース記事からの未来参照表現を利用することにより，ユーザーが日常的に行う日常的なタスクの枠組みの中で未来を予測するプロセスを改善できる可能性につながる．たとえば，日刊紙から取られた次の通常の未来関連の文章を分析すると，そのような文の信頼性を正しく推定し，将来の展開イベントの予測をサポートし，株式投資，企業管理，トレンド予測，リスク防止などにこの機能を適用できる．さらに，以前の研究で示されているように，ソーシャルネットワーキングサービス（SNS）の分析に使用され，自然災害や病気の発生を軽減が可能になる．

「年」,「時間」,「明日」などの時間関連の表現を利用する技術は，将来関連するデータや関連性の高いドキュメントを抽出するために利用されてきた．同様に，広くアクセス可能なコンテンツで発生するデータを利用して，イベントの将来の展開を推定することが役立つことが確認されている．残念ながら，上記で引用したすべての研究は明示的な未来関連の表現を対象にしているが，より洗練された暗黙のパターンを適用したものはない．したがって，このようなパターンを利用する技術は，新しい観点から未来を予測する問題を軽減し，将来のデータ抽出の一般的な研究に大きく貢献することが期待できる．

本研究では，単純な時間関連の表現や時系列に配列されたデータからの情報検索よりも洗練されたアプローチを適用することにより，将来の出来事について実際の生活の中で予測を行うための支援方法を開発することである．未来を参照するすべての可能な文型の自動抽出のための方法を提案し評価を行う．意味役割情報と形態素情報とニュースドメイン毎の特徴語の組み合わせを使用して一般化された将来の参照文から抽出されたばらばらの要素を持つ単語，フレーズ，より洗練された構造などのパターンを適用し，中島ら [10] が課題に残した将来言及文分類の精度向上を目指し，ニュースドメイン毎の将来言及文における特徴語の定義，ニュースドメイン毎の分類モデル生成を行い検証を行う．

本論文は，3 章では本研究で用いる将来言及文について，4 章では意味役割情報と形態素情報および未来語による形態情報ラベル，5 章で将来言及文分類モデル生成について説明し，6 章では実験設定と実験結果を示し，7 章では本実験の考察，最後の 8 章で本改善手法の結果と展望について述べる．

3 将来言及文

本研究で扱う将来言及文について説明する．

ある文が言及しているその時点よりも未来へ言及している文を将来言及文と定義する．将来へ言及する文には，明示的，暗示的ないくつかパターンが存在する．

将来言及文の例[1] を以下に示す．文 1,2,3 のように未来の時間情報 (年，月，日，来年，来月, 今後など) が明記されているパターン，文 4 に含まれている「方針」のように明らかに将来へ言及している単語を含む明示的に将来に言及している文である．文 5 は暗示的なパターンの文で未来の時間情報，将来を言及する単語が含んでいないが将来リストラされるということが言及されている．文 6 は，主節は過去形だが言及している内容は未来に言及している文であり全体で意図するのは未来へ言及している文である．

1. 三菱化学では 2025 年前後に Li イオン 2 次電池を搭載した 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。

公表される将来に言及している文は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はない限り，専門家または，その情報に精通している記者，有識者の見立てで書かれている．正当なニュース記事に関しては，数回のチェック機構を設けている．それらを考慮すると，将来に言及している文は社会情勢や過去の事実などの背景，専門家の調査結果や見立て，研究動向, 現在の事実など，素人では調べ上げるには困難なほどの知識的背景と事実に基づいて表現されていると考えられる．

将来言及文を用いた未来動向予測は中島ら [13] が，言論責任保証協会主催の先見力検定[2] の結果で示されている人間が 1 年間あらゆる情報を収集し予測するよりも，将来言及文 17 文〜30 文を用いた方が予測精度が高いことを明らかにしている．

我々は，予測支援文を収集する範囲を Web に拡張し，さらに，ニュースドメインの科学技術，経済，国際に対応する将来言及文分類モデルを生成し，未来動向予測に有用な将来言及文を獲得する．

4 意味役割情報と形態素情報および未来語による形態情報ラベル

提案手法は，1 つの日本語で表された文を意味役割情報，形態素情報と未来語の形態情報ラベルを生成し，その形態情報ラベルを学習させて将来言及文分類モデルを生成する．本章では，形態情報ラベルを構成する要素について述べる．

4.1 意味役割情報

意味役割情報は，文章の述語と項の関係を表した情報である．下記に示すように，入力文に対して述語項構造解析を行い，その後，述語の語義を同定し，係り関係にある項の意味役割を付与を行う．

[1] 日経 XTECH, 朝日 Digital，NHKNewsWeb，産経新聞

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

(入力文) 昨日母が私にお弁当を作った。

(出力) [昨日] 時間 (点)

[母が] 動作主

[私に] 着点 (人)

[お弁当を] 対象 (生成物)

[作った] 状態変化あり-生成・消滅-生成（物理）-生成-

文の述語に対して，概念を付与し，述語にかかる係り元 (項) に意味的な関係を付与する．例えば「母が」は「作った」に対して「動作主」という意味役割であることを示している．意味役割は ASA(Argument Structure Analyzer) [3] で扱う 72 種類で意味役割の一覧は岡山大学竹内研究室が作成した述語項構造シソーラス [4]を用いている．

本研究で対象とする文の解析と意味役割情報の付与には ASA を用いて意味役割情報を得る [11, 12]．ASA は意味役割の他に，かかり先 (Link) や時制 (Tense), 形態素 (Morphemes) などの情報を提供する (Figure: 1).

ID	0	**ID**	3
Surface	母が	**Surface**	作る。
Link	3	**Link**	-1
CType	elem	**CType**	verb
Main	母	**Main**	作る
Part	が	**Similar**	25
Category	["人"]	**Semantic**	状態変化あり-生成・消滅-生成（物理）-生成-
Semrole	動作主	**Voice**	ACTIVE
Similar	7	**Tense**	PRESENT
Tense	PRESENT	**Sentelem**	PREDICATE
Arg	["Arg0"]	**Polarity**	AFFIRMATIVE
Frames	["3-verb"]	**Mood**	INDICATIVE
Morphemes	0 母 ハハ 母 名詞,一般 O 1 が ガ が 助詞,格助詞,一般 O	**Frames**	["0-動作主-Arg0","1-elem","2-対象（生成物）-Arg1"]
		Morphemes	0 作る ツクル 作る 動詞,自立 基本形 O 1 。 。 。 記号,句点 O

Figure 1: 「母が」と「作った」に対して ASA を実行して得られる情報例

4.2 品詞情報

主に意味役割情報を用いるが，ASA が持たない単語や複合語に対応するために単語の品詞情報を付与する．品詞情報は，オープンソース形態素解析エンジン MeCab[5]を使用する．

例えば,「第一に、Ｓｏｃｉｅｔｙ５．０の実装です。」で使われている，“Ｓｏｃｉｅｔｙ５．０”は一つの単語として扱いたいが，ASA の結果はＳｏｃｉｅｔｙ (名詞), ５．０ (名詞) と別れてしまう．このような場合一つの単語として“Ｓｏｃｉｅｔｙ５．０”に名詞を付与する複合語処理を行う．

4.3 未来語

将来言及文の中には，明らかに未来を指し示す「見通し」「予定」などの語を含むパターンが存在する．それらの文を確実に取得するために，未来語を定義する．未来語は将来言及文とその他の文の頻出度と比較し，将来言及文に特に頻出する語とする．中島 [14] で将来言及文の表層的な特徴を「未来を示す時間情報」,「述語動詞」,「頻繁に使われる語彙」について調査し,「頻繁に使われる語彙」は疎であるが，特に頻繁に使われる語彙 31 語を未来語と定義をした．

本研究では，中島 [14] が定義した未来語の条件を改良し，科学技術，国際，経済それぞれの未来語，また，共通の未来語として定義する．

将来に言及している文を FRS(Future Reference Sentences) とその他の文を NFRS(Non-Future Reference Sentences) と表現する. 先ほど記述した通り，未来語とは，将来言及文と非将来言及文の語彙出現頻度数の差から，将来言及文の特徴となる語彙である．

未来語の抽出は，過去の記事から得た将来言及文と，その他の文を用いる．これらの文は，WEB 上の科学技術，経済，国際各分野のニュース記事からそれぞれ，将来言及文 500 文，その他の文 500 文　各分野合計 1000 文を収集した．これらの文は，信頼性の観点から Google 検索においてニュースに分類されるものに限った．将来言及文か他の文かの判断は将来言及文についての知識がある 3 名で行い，3 名の意見がすべて一致したものを将来言及文，その他の文としている．

未来語抽出の条件を双方に出現する差を考慮するために直接確率計算のしきい値を 3.0%以下に設定し，その他の文に出現する頻度を 0〜3 に限定する. この数値は，著者らが直接確率の値とそれぞれの頻度と単語を確認して決定した．ある単語がその他の文に出現する頻度に対して将来言及文に出現する頻度の条件を Table 1 のように設定する．ただし，m をその他の文に出現する頻度，n を将来言及文に出現する頻度とする．

Table 1: 未来語抽出のための単語出現頻度の条件

	ある単語の出現頻度 (回)			
その他の文	$m = 0$	$m = 1$	$m = 2$	$m = 3$
将来言及文	$n \geq 5$	$n \geq 7$	$n \geq 14$	$n \geq 21$

設定条件を適用し，科学技術，国際，経済分野と 3 分野の共通ごとに抽出した単語を以下に示す．

科学技術分野

順次　達する　進む　将来的　発展　実用化　商品化　予想　加速　めど　将来　発売　完了　一気に　見える　知見　適用　来月　来年　来週　方針　計画　今後　目指す　可能性　見通し　そう　予定　さらに　検討　始まる　発表　見込み　見込む　予測　目標　展開

経済分野

恐れ　決定　年内　意向　想定　継続　表明　少なくとも　出る　実現　続ける　難航　再開　協議　示唆　見方　影響　来月　来年　来週　方針　計画　今後　目指す　可能性　見通し　そう　予定　さらに　検討　始まる　発表　見込み　見込む　予測　目標　展開

国際分野

取りやめる　今週　訪れる　移動　撤去　中止　延期　来週　考え　火種　難航　来月　来年　来週　方針　計画　今後　目指す　可能性　見通し　そう　予定　さらに　検討　始まる　発表　再開　協議　示唆　見方　影響

[3]http://www.cl.cs.okayama-u.ac.jp/study/project/asa/

[4]http://pth.cl.cs.okayama-u.ac.jp/

[5]https://taku910.github.io/mecab/

3 分野共通

来月　来年　来週　方針　計画　今後　目指す　可能性
見通し　そう　予定　さらに　検討　始まる　発表

未来語は科学技術分野に 37 個，経済分野に 37 個，国際分野に 31 個，3 分野共通に 15 個が抽出された．

4.4　形態情報付与

文章を意味役割情報，形態素情報，未来語を用いた形態情報ラベルについて説明する．1 文ごとに以下の優先順位に従い，形態情報ラベルを生成する．1 の未来語ラベルは，未来語群に該当する単語に “未来語” と統一したラベルとする．2,3,4 については，Figure 1 で示した複数の情報から **Semrole, Semantic, Category** の 3 つの情報を以下の優先順位に従いラベルとする．

1. 未来語
2. 意味役割情報 (Semrole)
3. 述語意味論情報 (Semantic)
4. カテゴリー情報 (Category)
5. 複合語処理後の品詞

文を形態情報ラベル付与の例を以下に示す．

例 1：原文　２０２３年３月末までの製品化を目標としている。

形態情報ラベル　名詞 時間 対象 未来語 状態変化あり

例 2：原文　１６年度末からの納入を目指し、最終的に４２機を配備する。

形態情報ラベル　名詞 モノ 対象 未来語 名詞 助詞 対象 状態変化あり

5　将来言及文分類モデル生成

将来言及文分類モデルは，形態情報ラベルを学習データとすることで，1〜6 ラベルで構成する形態パターンを獲得し，将来言及文分類モデルを生成する．まず，1〜6 ラベルで構成する形態パターンについて説明する．次に，将来言及文分類モデルについて説明する．

5.1　形態パターン

形態情報ラベルで表した各ラベル要素を 1〜6 ラベルのすべての組み合わせで表される．

この変換した情報を元に，SPEC(Sentence Pattern Extraction and analysis arChi-tecture) [15] を用いて形態パターンの作成を行う．SPEC はまず，すべての文の要素から順番を考慮した組み合わせを生成する．生成するパターンの要素を n とすると，$1 \leq k \leq n$ を満たす整数 k 個の組み合わせグループが存在することになる．また，ワイルドカードを意味するアスタリスク「*」は，0 個以上のラベルを意味する．

このようにして学習データから抽出されたパターンリストを学習することで分類モデルの生成を行う．また，このパターンリストは，与えられたデータに出現するパターンを抽出し，その重み計算を行う．この重み値が，しきい値以上なら将来言及文，未満なら非将来言及文と分類することになる．SPEC では重み計算方法が 14 種類あり，そのうち交差検定の結果で最もよい精度の結果を採用する．しきい値は，分類モデルの評価時に適合率，再現率および F 値の値を観察し最もバランスの良いポイントをしきい値に設定する．

例えば，「太郎は、花子に花を贈るだろう。」の形態情報ラベルは Section:4 で示した手法を適用すると次のようにラベリングされる．

ラベリング: [動作主][着点][対象][状態変化あり]

この場合，意味役割ラベル要素は 4 個であり，全ての要素の組み合わせが形態パターンとできる．形態パターンの例を以下に示す．

形態パターン例:

[動作主]
[動作主][着点]
[着点][対象]
[対象]*[状態変化あり]
[動作主]*[状態変化あり]
[動作主][着点]*[状態変化あり]
[動作主]*[対象][状態変化あり]
[動作主][着点][対象][状態変化あり]

5.2　将来言及文分類モデル

文は意味役割ラベルに変換されたのち，SPEC を用いて形態パターンを獲得し学習を行う．正解データに将来言及文と不正解データにその他の文を入力として，将来言及文分類モデルを出力する．将来言及分類モデル機構の概要を Figure:2 に示す．

本研究では，将来言及文 500 文とその他の文 500 文合計 1000 文の入力で学習し分類モデルを生成する．データ数に関しては，中島ら [13] が SPEC の学習において 260 文のデータで実験をしており，データ数が少なくある程度の精度が得られることを示している．

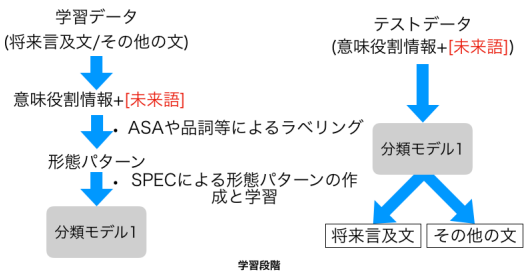

Figure 2: 将来言及文分類モデル機構の概要

6 実験

本章では，文を意味役割情報，品詞情報および未来語を用いた形態情報ラベルで表現し，形態パターンを用いた将来言及文分類モデル生成とその評価について実験を通して有用性の検証を行う．本研究の目的である，汎用型を実現するために，ニュース記事分類のうち，科学技術，国際，経済各分野において実験を行う．

6.1 実験設定

本実験で，扱うデータを文の構成が比較的整っている web 上のニュース記事に限定する．個人が書くブログや twitter などの文章は除外する．各分野の区別は，web 上のニュース記事に付与されている分野のタグを利用する．

将来言及文分類器生成のための訓練データの正解データ（将来言及文），不正解データ（その他の文）は，あらかじめ将来言及文についての知識がある 3 人で行い，3 人の意見が一致したものを採用する．

訓練データには，各分野の将来言及文 500 文，その他の文 500 文の合計 1000 文とする．

形態パターン，および分類モデル生成には SPEC を利用する．分類モデルの評価は 10 分割交差検定で評価を行う．また，精度検証のために機械学習に一般的に利用されている SVM(Support Vector Machine) [16] の結果と比較を行う．

評価指標は再現率と適合率，および，その調和平均である F 値を用いる．再現率とは，実際に将来言及文である文のうち，将来言及文であると予測された文の割合である．また，適合率とは，将来言及文と予測した文のうち，実際に将来言及文である文の割合である．

6.2 実験結果

実験結果を分野別に適合率，再現率，F 値で Table 2 に示す．SVM は Python のオープンソース機械学習ライブラリである scikit–learn[6]を用いて実験を行った．

Table 2: 科学技術，経済，国際分野ごとの将来言及文分類モデル評価結果 : SPEC と SVM の比較

	適合率		再現率		F 値 (調和平均)	
分野	SPEC	SVM	SPEC	SVM	SPEC	SVM
科学技術	0.990	0.851	0.930	0.65	0.960	0.737
経済	0.960	0.828	0.910	0.712	0.930	0.766
国際	0.990	0.853	0.860	0.766	0.920	0.807

7 ディスカッション

本実験結果について考察する．本実験で使用したデータは，将来言及文について知識のある 3 人によって慎重に選び使用した．

従来研究で利用されている，時系列情報および時間情報，表層的な未来時制を表す語（例えば英語では will や be goin to など）に依存することなく，将来言及文の意味役割情報と品詞情報および新たに定義した未来語を組み合わせた形態パターンを用いることで，明示的な将来言及文の他に暗示的な将来言及文も獲得可能な手法を提案した．

Table 2 に示される値のほとんどが本手法で採用した SPEC による結果が SVM による結果よりも高い値であった．SPEC の特徴の一つであるデータセットが小さくてもある程度の精度が得られるという理由が貢献した結果となったと考える．また，SPEC は SVM の結果と比較して，ばらつきが少ないこともわかる．

分野別の F 値（適合率と再現率の調和平均）においては，科学分野 0.960, 国際分野 0.920，経済分野 0.930 と 0.9 以上の高い評価値の分類モデルであった．国際分野の再現率が 0.860 ではあるものの，概ね網羅的に将来言及文が獲得できる範囲であると考える．

また，中島ら [12] が行った意味役割情報と品詞情報を用いた形態パターンを用いた分類モデルの精度は F 値 0.76 であり，未来語を形態パターンの要素して加えることで精度の改善が確認できた．彼らの実験結果でエラー文となった文からランダムに 20 文を選び，本手法を適用し比較を試みた．SPEC では分類モデルで文を分類する際に，正解度に相当する値が計算される．本実験においては，将来言及文らしさの値と置き換えることができ，その値を比較すると約 7 割の文の将来言及文らしさの値が改善された．その値によって，正しく分類された文の例[7]を以下に示す．未来時間情報などの表層的な文の末尾には “明示的", 暗示的に表現されている未来言及文の末尾には “暗示的"と示す．文中の個人名はアルファベット昇順に置き換えている．

将来言及文として分類できた文の例：

- 政権発足後、A 大統領は評判の悪いグアンタナモ収容所の閉鎖を命じ、イラク撤退の道筋をつける半面、アフガニスタンには米軍増派の決断を下した。(暗示的)
- B 首相は２３日、C 米大統領との初の日米首脳会談に臨む。(明示的)

その他として分類できた文の例：

- ノーベル平和賞授賞式の演説で D 米大統領が武力行使の意義を説いたのも、弱いイメージを持たれたくないためだろう。
- E 大統領は大恐慌で疲弊した米国を救った「ニューディール政策」で有名ですが、政策実施に必要な１５件の法案を就任後１００日間に議会で成立させました。

8 まとめと展望

将来言及文を分類するための分類モデルの精度向上を目的とし，意味役割情報と品詞情報に新しく未来動向を指し示す未来語を定義し，将来言及文を意味役割情報，品詞方法と未来語の形態パターンを考案し，また，ニュースドメインのうち科学技術，国際および経済のドメインごとにデータを学習することにより将来言及文の分類モデルを生成し，評価を行った．本手法による分類モデルの精度は，F 値で科学分野 0.960, 経済分野 0.930，国際分

[6]http://scikit-learn.org/stable/

[7]毎日新聞データ集 2009

野 0.920 の結果が得られ，精度改善が確認できた．また，先行研究でエラー文となった文に対しても 7 割の改善が確認できた．今後，本実験で生成した分類モデルを用いて，Web 上のニュース記事のほか政府官邸，各省庁の政策，および綜合研究所の未来展望など専門家がまとめた記事に含まれる将来言及文の分類実験を行うなど，さらに，本分類器の精度を確認する必要がある．汎用性に関しては，ドメイン毎に分類モデルを生成することで，精度のより高い汎用型分類モデルが可能であることが明らかになった．政策などのドメインを加増や各ドメインを自動選択する機能を加え適切な分類モデルを選択することで実現を目指す．

9 謝辞

本研究は JSPS 科研費 17K00324 の助成を受けたものである．

References

[1] K.Shirata, H. Tsuda. (2018). 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.

[2] H.Maekawa, T.Nakahara, K.Okada, et al. (2013). 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.

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

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

[5] Nakajima, Y., Ptaszynski, M., Honma, H., & Masui, F. (2014). Investigation of future reference expressions in trend information. In AAAI spring symposium series big data becomes personal: Knowledge into meaning.

[6] Nie, A., Choi, J. D., Shepard, J., & Wolff, P. (2015). Computational exploration to linguistic structures of future: Classification and categorization. In Proceedings of the 2015 conference of the North American Chapter of the Association for computational linguistics: Student research workshop (pp. 168–173).

[7] Al-Hajj, M., & Sabra, A. (2018). Automatic identification of Arabic expressions related to future events in Lebanon ' s economy. International Journal of Science and Research (IJSR), 7(4), 1656–1660.

[8] Yarrabelly, N., & Karlapalem, K. (2018). Extracting predictive statements with their scope from news articles. In Twelfth International AAAI conference on web and social media.

[9] Hurriyetoglu, A., Oostdijk, N., & van den Bosch, A. (2018). Estimating time to event of future events based on linguistic cues on twitter. In Intelligent natural language processing: Trends and applications (pp. 67–97). Cham: Springer.

[10] Y.Nakajima, M. Ptaszynski, F. Masui, H. Hirotoshi. (2016). 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.

[11] 竹内孔一，森本真依子. (2009). 動詞項構造シソーラスに基づく動詞語義ならびに意味役割付与データの構築. 言語理解とコミュニケーション研究会, NLC2009-9 pp.13–18.

[12] 原 靖弘, 竹内 孔一. (2016). 係り元の末尾表現に着目した Hierarchical Tag Context Tree を利用した日本語意味役割付与システムの構築, 情報処理学会論文誌 Vol.57 No.7 1611–1626.

[13] Y.Nakajima, M. Ptaszynski, F. Masui, H. Hirotoshi. (2017). 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.

[14] 中島陽子. (2017). 未来イベント予測のための将来言及文における特徴語の調査. 釧路工業高等専門学校紀要, vol.51, pp64–68.

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

[16] Cortes, C., & Vapnik, V. (1995). Support vector machine. Machine learning, vol.20(3), pp273–297.