

令和 2 年 7 月 6 日現在

機関番号：50103

研究種目：基盤研究(C)（一般）

研究期間：2017～2019

課題番号：17K00324

研究課題名（和文）テキストの形態パターン情報を活用した汎用型未来動向予測支援システムの開発

研究課題名（英文）Development of a versatile future trend prediction support system using morphosemantic patterns information of text data

研究代表者

中島 陽子（Nakajima, Yoko）

釧路工業高等専門学校・創造工学科・准教授

研究者番号：20217730

交付決定額（研究期間全体）：（直接経費） 3,500,000円

研究成果の概要（和文）：本研究では、ニュース記事に含まれる将来に言及している文を自動的に抽出し、それらの文を用いた未来動向予測支援システムの構築を目的とした。研究成果として、限定的な語彙を手掛かりとしない、潜在的な将来に言及する文の抽出・分類する技術、組合せベースの学習機能により有用な少量の学習データで分類モデル生成が可能であることを明らかにし、また、対象分野の専門知識を用いることなく将来イベントに言及する文を用いた未来予測を支援する技術を開発した。本システムの有用性は過去のニュース記事から将来に言及している文を抽出し、1-2年後のイベントを予測する実験により確認した。

研究成果の学術的意義や社会的意義

本研究は、従来手法の手がかり表現のような限定的な語を用いることなく、文中に混在する潜在的意味を持つような文に対しても応用可能であり有用な文を取得する手法として貢献できると考える。また、従来手法とは違う手法で未来予測の可能性を示したことは、他分野への発展が期待できる。本システムは、専門的知識を持たずとも近年1,2年の将来言及するデータをもとに企業・政策活動などにおいて、中・長期戦略などで必要となる先を見通すための支援に応用が可能であり、さらには、社会技術イノベーション創出に貢献できると考える。

研究成果の概要（英文）：In this research, we developed a versatile future trend prediction support system using future reference sentences extracted automatically from news articles.

To achieve that, we firstly developed a method to classify and extract implicit and explicit future reference sentences without the use of simple keywords, and confirmed that it is possible to generate a classification model with a small amount of useful learning data by using a Language Combinatorics-based learning method. Next, we applied automatic future reference extraction to develop a method to support future trend prediction using only future reference sentences, with no need for expert knowledge. The usefulness of the system as a whole was verified in experiments in which sentences referring to the future were extracted from news articles and related to them events were predicted in 1-2 years later.

研究分野：自然言語処理

キーワード：自然言語処理 意味役割 将来言及文 未来動向予測支援 形態パターン 機械学習 分類モデル

## 1. 研究開始当初の背景

近年、株価変動や社会情勢動向などを対象とした精度の高い未来予測の実現要求が高まっている。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ることから、これら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世界的に注目されつつある。具体的には、類似尺度から得た因果的推論を使用した将来イベント予測システムの研究、テキスト情報のクラスタリングによるイベントの生起確率計算方法を提案した研究、Twitter データの機械学習によってウイルス拡散予測を試みた研究などがある。

これら未来予測システムの多くはテキスト中の時間参照情報を用いた統計処理や機械学習による予測手法を採用している。しかしながら、これらの手法には、それぞれで、問題固有の統計処理技法や、大量の学習データ処理が必要であり、予測対象分野ごとに異なる専門的観点からのヒューリスティクスを要するため、汎用性に問題があった。そこで、実用性と汎用性を兼ね備えた未来予測手法を提案するに至った。

## 2. 研究の目的

本研究は、文を構成する形態パターンから抽出される将来言及文を活用した新たな未来イベント予測手法の実現に挑戦する。本研究の目的は、(1)対象分野の専門知識を用いることなく、(2)少量の学習データで潜在的な将来言及文を抽出・分類可能、(3)汎用性が高い未来動向予測支援システムを開発することである。

## 3. 研究の方法

本研究の研究方法は、主にニュース記事や政府が web 上で公開するデータなどを用い、次に示すように段階的に行った。

(1)意味役割付与機構の構築、(2)形態パターン導出機構の構築、(3)将来言及形態パターン(MoPs)学習機構の構築、(4)複数コーパスからの MoPs 抽出実験、(5)将来言及文取得のための分類モデルの構築、(6)予測エンジンの構築、(7)未来動向予測システムの構築、(8)未来動向予測実験と評価、(9)将来イベント予測の実問題への応用の検討。

各段階の後でも、改良が必要な場合が生じた際には、適宜、改善しながら研究を行った。

## 4. 研究成果

本システムの概要を図 1 に示す。形態パターンを用いた将来言及文の抽出と将来言及文の有用性について、将来言及文分類モデルの構築、予測エンジン部の構築、未来動向予測支援システムの構築と評価に分けてそれぞれの研究成果について報告する。

### (1) 形態パターンを用いた将来言及文の抽出と将来言及文の有用性

新聞記事と web ニュース記事から将来言及文とその他の文を各 500 文取得し、形態素情報と文脈解析に用いられる意味役割を併用した形態パターン(Morphosemantic Patterns: MoPs)を生成し、

教師あり機械学習を用い将来言及文を分類するための分類モデルを生成した。さらに、精度の向上を目指し、将来言及文の特徴となる語を調査し、それらを未来語として定義し、前述の MoPs に未来語を素性として加え、分類モデルを生成し比較を行なった。また、ニュース記事は国際、科学技術、政治などドメインに分類されていることを利用し、科学技術の記事に注目し実験を行なった。

以上の手法により抽出した将来言及文を用いた未来イベント予測を行う際に、将来言及文が有効であるか否かを検証するためのプロトタイプ予測エンジンを設計し実験を行なった。予測エンジンは、将来言及文を入力とし、機械学習により正解データを分類する正解分類モデルを生成し、この分類モデル

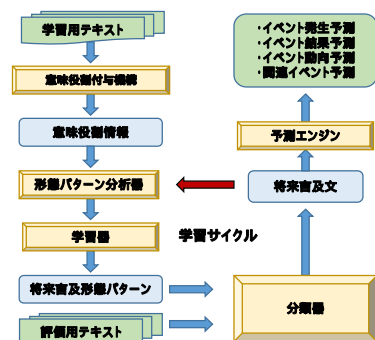


図 1: 未来動向予測システム

に問題文を入力とし、2 個以上の選択肢の中から正解を選び解答する。将来イベントを予測する設問 7 問を人間と本手法により実験した結果、人間による予測実験の正解率は 30%程度に対し、プロトタイプ実験では、50%以上の結果が得られ、将来イベントを予測するための情報として、将来言及文の有用性を確認できた。研究成果については[3,4,5]にて発表している。

## (2) 将来言及文分類モデルの構築

未来動向予測支援システムの中核部である将来言及文獲得のための分類モデルの精度向上と汎用性の強化を実現し最適な分類モデルの構築、および、汎用性を備えた予測エンジンの開発を行った。

将来言及文獲得を行う分類モデルの精度向上には、将来言及文に使われる特徴的な語を未来語と定義し、形態パターンの要素(意味役割:動作主、動作など)に「未来語」を新たに追加し分類モデルを生成した。未来語は分野ごとに将来言及文 1000 文を用いて決定した。未来語を含まない潜在的な未来言及文も存在するため、要素に未来語を考慮する学習と未来語を意味役割に置き換え学習する二段階学習構造による分類モデルの生成を試みた。

各分類モデルの評価には、新たに収集したデータ 1200 文を用いて行い、H29 年度の実験結果よりも約 11 ポイントの向上を確認した。

分類モデルの汎用性強化と最適化実装は、ニュース記事の分野分類を参考に、科学技術分野の分類モデルに加え、国際、経済、政策の分類モデルを生成し、ニュースドメイン毎に適切な分類モデルを切り替え可能とする機能を追加した。さらに各分野と全分野共通の未来語を定義し、汎用性を考慮した。研究成果については[6,7]にて発表している。

## (3) 予測エンジン部の構築

予測エンジン部は、実世界の問題に対応するために予測したい未来動向が[起きるか/起きないか/どちらともいえない]の回答を得る手法として、単語の極性が[ポジティブ/ネガティブ]であるかを表す日本語評価極性を用いた手法を考案し実験を行った。予測エンジンの精度は、プロトタイプ予測エンジンよりも分野混合動向予測問題の予測結果は 23.9 ポイント、分野別の予測結果は 12.5 ポイント高い結果が得られた。

しかし、単語の極性が[ポジティブ/ネガティブ]であるかを表す日本語評価極性を用いた手法では辞書依存が原因で日本語評価極性辞書にない単語が頻出すると判定できないという問題を解決するために、分野ごとに頻出単語を辞書補完し、極性判定を定式化し極性分類精度の改善を行った。研究成果については、[8]にて発表している。

## (4) 未来動向予測支援システムの構築と評価

将来言及文抽出精度、予測エンジンの予測精度を向上させるため、各過程において検討・精査、およびアルゴリズムの改善を行い、未来動向予測システムの構築および、実世界における予測問題とデータを用いシステムの実用性の検証を行った。

意味役割付与機構、形態パターン導出機構、将来言及形態パターン導出機構、将来言及文分類モデル、予測エンジンを組合せ、将来動向予測システムを構築した。

本システムの評価は、既存の「先見力検定」で出題された問題、結果が判明している実世界のイベントから我々が作成した未来予測問題を用い Web ニュース記事を入力として実験を行った。全 15 問の平均正解率は約 73%であり人間の平均予測正解率 30%を大きく上回る結果を得た。研究成果については [8,9]にて発表している。

本研究が目的とする、対象分野の専門知識を用いることなく、文を構成する形態パターンを用い少量の学習データで潜在的な将来言及文を抽出・分類できる、汎用性を備えた未来動向予測支援システムの開発は計画通り遂行できた。

### < 引用文献 >

[1] Y. Nakajima, M. Ptaszynski, H. Honma, F. Masui, A Method for Extraction of Future Reference Sentences Based on Semantic Role Label, *IEICE Trans. Information and Systems*, Vol.E99D, No.2, 2016, pp.514-524 (査読有)

- [2] Yoko Nakajima, Michal Ptaszynski, Hirotoishi Honma, Fumito Masui, Automatic Extraction of Future References from News Using Morphosemantic Patterns with Application to Future Trend Prediction, *ACM Summer Newsletter AI Matters*, Vol.2, No.4, 2016 , pp.13-15 (査読有)
- [3] Yoko Nakajima, Michal Ptaszynski, Hirotoishi Honma, Fumito Masui. A Prototype Method for Future Event Prediction Based on Future Reference Sentence Extraction, *Linguistic and Cognitive Approaches To Dialogue Agents (LACATODA 2017)*, 2017, pp.42-49
- [4] Yoko Nakajima, Michal Ptaszynski, Hirotoishi Honma, Fumito Masui, Future Reference Sentence Extraction in Support of Future Event Prediction, *International Journal of Computational Linguistics Research*, Vol.9, No.1, 2018, pp.27-41
- [5] Michal Ptaszynski, Fumito Masui, Yoko Nakajima, Yasutomo Kimura, Rafal Rzepka, Kenji Araki, A Method for Detecting Harmful Entries on Informal School Websites Using Morphosemantic Patterns, *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol.21, No.7, 2017, pp.1189-1201 (査読有)
- [6] 中島陽子, 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, *釧路高専紀要*, Vol.51, 2017, pp.64-68
- [7] 中島陽子, 本間宏利, Akmal Hakim, Michal Ptaszynski, 榎井文人, 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, *釧路高専紀要*, Vol.53, 2020, pp.37-42
- [8] Yoko Nakajima, Michal Ptaszynski, Hirotoishi 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
- [9] Yoko Nakajima, Michal Ptaszynski, Fumito Masui, Hirotoishi Honma, Future Prediction with Automatically Extracted Morphosemantic Patterns, *Cognitive Systems Research*, Vol.59, 2020, pp.37-62 (査読有)

## 5. 主な発表論文等

〔雑誌論文〕 計16件（うち査読付論文 13件 / うち国際共著 7件 / うちオープンアクセス 5件）

1. 著者名 Yoko Nakajima, Michal Ptaszynski, Hirotooshi Honma, Fumito Masui	4. 巻 vol.9, No.1
2. 論文標題 Future Reference Sentence Extraction in Support of Future Event Prediction	5. 発行年 2018年
3. 雑誌名 International Journal of Computational Linguistics Research	6. 最初と最後の頁 27-41
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Michal Ptaszynski, Fumito Masui, Yoko Nakajima, Yasutomo Kimura, Rafal Rzepka, Kenji Araki	4. 巻 Vol.21, No.7
2. 論文標題 A Method for Detection of Harmful Entries on Informal School Websites with Morphosemantic Patterns	5. 発行年 2017年
3. 雑誌名 Journal of Advanced Computational Intelligence and Intelligent Informatics	6. 最初と最後の頁 1198-1201
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 中島陽子	4. 巻 第51号
2. 論文標題 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	5. 発行年 2017年
3. 雑誌名 釧路工業高等専門学校紀要	6. 最初と最後の頁 64-68
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Iwabuchi Rio, Nakajima Yoko, Honma Hirotooshi, Aoshima Haruka, Kobayashi Akio, Akiba Tomoyoshi, Masuyama Shigeru	4. 巻 -
2. 論文標題 Proposal of recommender system based on user evaluation and cosmetic ingredients	5. 発行年 2017年
3. 雑誌名 Advanced Informatics, Concept, Theory, and Applications on IEEE Xplore Digital Library	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/ICAICTA.2017.8090967	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Nakajima Yoko, Ptaszynski Michal, Masui Fumito, Honma Hirotooshi	4. 巻 59
2. 論文標題 Future prediction with automatically extracted morphosemantic patterns	5. 発行年 2020年
3. 雑誌名 Cognitive Systems Research	6. 最初と最後の頁 37～62
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.cogsys.2019.09.004	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 中島陽子, 本間宏利, Akmal Hakim, Michal Ptaszynski, 榎井文人	4. 巻 53
2. 論文標題 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	5. 発行年 2020年
3. 雑誌名 釧路高専紀要	6. 最初と最後の頁 37～42
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計21件 (うち招待講演 0件 / うち国際学会 11件)

1. 発表者名 Yoko Nakajima, Akiko Saito, Rio Iwabuchi, Hirotooshi Honma, Tomoyoshi Akiba
2. 発表標題 Identification of Harmful Ingredients of Cosmetics using User Comments
3. 学会等名 JSST Annual International Conference on Simulation Technology (国際学会)
4. 発表年 2018年

1. 発表者名 Michal Ptaszynski, 榎井文人, 中島陽子, 早川博, 齊藤剛彦, 宮森保紀
2. 発表標題 Comparison of Machine Learning Techniques for Classification of Information Types on Twitter
3. 学会等名 言語処理学会第25回年次大会
4. 発表年 2019年

1. 発表者名 Yoko Nakajima, Michal Ptaszynski, Hirotooshi Honma, Fumito Masui
2. 発表標題 A Prototype Method for Future Event Prediction Based on Future Reference Sentence Extraction
3. 学会等名 Linguistic and Cognitive Approaches To Dialogue Agents (LACATODA 2017) on IJCAI (国際学会)
4. 発表年 2017年

1. 発表者名 Rio Iwabuchi, Yoko Nakajima, Hirotooshi Honma, Haruka Aoshima, Akio Kobayashi, Tomoyoshi Akiba, Shigeru Masuyama
2. 発表標題 Proposal of Recommender System Based on User Evaluation and Cosmetic Ingredients
3. 学会等名 Advanced Informatics, Concept, Theory, and Applications (ICAICTA) (国際学会)
4. 発表年 2017年

1. 発表者名 Michal Ptaszynski, Yuka Ito, Karol Nowakowski, Hirotooshi Honma, Yoko Nakajima, Fumito Masui
2. 発表標題 Combining Multiple Dictionaries to Improve Tokenization of Ainu Language
3. 学会等名 The 31st Annual Conference of the Japanese Society for Artificial Intelligence
4. 発表年 2017年

1. 発表者名 猫本隆哲, 澤山大陸, 中島陽子, 本間宏利, Michal Ptaszynski, 榎井文人
2. 発表標題 ヒューリスティック手法による隠喩判定システムの精度改善
3. 学会等名 言語処理学会第24回年次大会
4. 発表年 2018年

1. 発表者名 Yoko Nakajima, Michal Ptaszynski, Hirotoishi Honma, Fumito Masui,
2. 発表標題 A Proposal of Prediction Method Using Word Polarity Information for Future Event Prediction Support System
3. 学会等名 Advanced Informatics, Concept, Theory, and Applications (ICAICTA) (国際学会)
4. 発表年 2019年

1. 発表者名 Yoko Nakajima, Hirotoishi Honma, Tomoyoshi Akiba, Shigeru Masuyama
2. 発表標題 Recommender System Based on User Evaluations and Cosmetic Ingredients,
3. 学会等名 The 4th International Conference on Information Technology (InCIT) (国際学会)
4. 発表年 2019年

1. 発表者名 Yuka Takahashi, Tsendsuren Urangoo, Yuto Tamori, Takumi Maruyama, Yoko Nakajima, Hirotoishi Honma, Kazuhide Yamamoto
2. 発表標題 Automatic Generation of Supplementary Conjunction Questions for Learners of Japanese
3. 学会等名 The 4th International Conference on Science of Technology Innovation (STI-Gigaku 2019) (国際学会)
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分 担 者	プタシンスキ ミハウ  (Ptaszynski Michal)  (60711504)	北見工業大学・工学部・准教授    (10106)	



## 6. 研究組織（つづき）

	氏名 (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分 担 者	榊井 文人  (Masui Fumito)  (80324549)	北見工業大学・工学部・教授    (10106)	
研究 分 担 者	本間 宏利  (Honma Hirotoshi)  (80249721)	釧路工業高等専門学校・創造工学科・准教授    (50103)	