

平成 30 年 5 月 31 日現在

機関番号：12601

研究種目：若手研究(B)

研究期間：2016～2017

課題番号：16K16109

研究課題名(和文) ソーシャルビッグデータ解析のための適応的言語解析

研究課題名(英文) Adaptive Natural Language Processing for Analyzing Social Big Data

研究代表者

吉永 直樹 (Yoshinaga, Naoki)

東京大学・生産技術研究所・准教授

研究者番号：90773961

交付決定額(研究期間全体)：(直接経費) 3,000,000円

研究成果の概要(和文)：モバイル端末を通して人々がマイクロブログにリアルタイムに発信する体験や意見はソーシャルビッグデータと呼ばれ、その有用性が注目されている。特に、新しい事物(固有表現)に関する情報はその分析価値が高いが、過去のデータを用いた静的な学習に基づく既存の自然言語処理技術ではこのような新情報を含むテキストの解析が難しい。

そこで本研究では、Wikipediaを利用した遠距離教師あり学習により多様な新固有表現をテキストストリームから動的に獲得する手法を開発した。また、獲得した知識を利用することを想定して、マイクロブログから大衆の価値観を獲得する手法や、機械翻訳、対話システムなどの研究も行った。

研究成果の概要(英文)：Experiences and opinions submitted by people in real time through mobile devices form so-called social big data, and their usefulness attracts attention. Although the information on emerging things (entities) is especially worth analyzing, it is difficult to analyze such text with new information by using existing natural language processing technology based on static machine learning with past text.

This study has therefore developed a method of acquiring various emerging entities from a text stream using distant supervision with Wikipedia. To utilize the acquired knowledge, we have also studied a method of acquiring values from microblogs, a machine translation system, and a dialogue system.

研究分野：Natural Language Processing

キーワード：自然言語処理 ソーシャル・ビッグデータ 知識獲得

## 1. 研究開始当初の背景

モバイル端末と Twitter に代表されるマイクロブログの普及により、一般の人々が自身の体験や意見を気軽に投稿・共有する時代となっている。人々が発信する情報は実世界のあらゆる時空間を網羅し、これら「ソーシャルビッグデータをリアルタイムで解析することで、世論の分析にとどまらず、竜巻などの局地的自然災害の検知、企業・自治体が提供する製品・施策の問題の発見、物流の投機的最適化など様々な形でより良い社会の実現に繋がることを期待できる。

ソーシャルビッグデータを解析するに際しては、新しい事物(人物、イベント、芸術作品など)の話題や既知の事物に関する新事実(自治体の新施策など)が重要な関心事となるが、既存の自然言語処理技術は解析に利用するモデルや補助的知識を事前に教師あり学習等で構築するため、これらの新情報を適切に解析することが難しい。例えば「グーグルが、これからはアルファベットで取引開始するって言う」という投稿を係り受け解析する際、アルファベットが企業であるという知識がなければ、「アルファベットで→取引開始する」ではなく、より多く出現する「アルファベットで→言う」と解釈する可能性が高い。一方、「アルファベットと認識できないフォントがすごい」という文を解析する際、アルファベットが企業であるという知識があると、逆に「アルファベット(=企業)と→認識できない」とは解析されづらくなる。

この問題を解決するには、人間のように同時期に発信される投稿と合わせて、臨機応変に(話題の変化を意識し、必要であれば知識を更新して)テキストの内容を読み解く必要がある。

## 2. 研究の目的

本研究では、ソーシャルビッグデータのようなソーシャルメディアストリームから、解析のための知識を陽に取り込み、自動更新して利用する適応的な言語解析技術の確立を目指し、ソーシャルメディアストリームから新固有表現(例:「アルファベット」)を獲得する手法を開発する。また、獲得した新固有表現を知識として活用することを念頭に、そのような新固有表現を含みやすい実世界テキストを入出力とする自然言語処理応用技術の開発も進める。

具体的に、ソーシャルメディアストリームからの新固有表現の獲得については、代表的なマイクロブログである Twitter から、多様な分野の新固有表現をオンラインで獲得する手法の確立を目指す。獲得の対象とする固有表現については、新聞など公共性を有するメディアでその出現が紹介されたり、Wikipedia に登録されたりするような公益性が高い固有表現を特に重視するが、それに限らず、個別

の商品など Twitter で話題になる固有表現を広く獲得できる手法を開発する。

また、実世界テキストを入出力とする自然言語処理応用技術としては、ソーシャルメディアの投稿を用いて多様な観点に基づき事物を順序付けする価値観獲得手法の開発と、機械翻訳や対話システムなど入力に新固有表現が頻出すると予想される自然言語処理応用技術の開発を行う。

## 3. 研究の方法

本研究では、ソーシャルビッグデータにおいて新事物を含むテキストを解析するのに必要となる新固有表現の動的な獲得手法の研究の開発に一貫して取り組んだ。以下で、この手法の開発および評価方法について詳しく述べる。

また、得られた新固有表現を利活用可能な実世界テキストを対象とした自然言語処理応用についても合わせて研究開発を行う。これらの応用技術の開発および評価方法についても、簡潔に説明を行う。

### (1) ソーシャルメディアストリームからの新固有表現の発見

リアルタイムで実世界イベントの話題が投稿される Twitter を対象として、投稿時点で Wikipedia に登録されていない新固有表現を、その場で自動的に発見して収集する手法の研究開発を行った。提案手法では、新固有表現が新たに発生して普及・認知される過程では固有表現の新規性を示唆する言及が多く観測されることに注目し、新固有表現(と既知固有表現)の出現文脈を学習データとして収集する。続いて、それらの投稿を用いて新固有表現を認識する系列ラベリングモデルを学習する。

教師あり学習では、一般に大量の学習データを要求されるが、データへの教師ラベルの付与は基本的に人手で行う必要があるためコストが大きい。また今回のタスクでは、人手でラベル付与を行う場合、将来的に固有表現の知識ベースが更新されたり、発見対象の新固有表現の性質や種類が変化したりすると、学習データを再構築する必要があると予想される。そのため学習データを自動生成する仕組みがあることが望ましい。この点を考慮し、既存の知識ベースを手がかりに擬似的な教師データを生成する遠教師あり学習を利用して擬似的な学習データを大量に自動生成して、このデータを用いて新固有表現をソーシャルメディアストリームからリアルタイムで検出するモデルを学習する。

前述の通り、新事物は出現初期の段階ではその新規性を示唆する表現で発信されやすい。例えば、以下は Twitter で映画「劇場版ポケットモンスター みんなの物語」を紹介する投稿であり、映画の新規性が示唆されている。

“『ポケモン』新作映画のタイトルが『劇場版ポケットモンスター みんなの物語』に決定！ 予告編映像も公開  
<https://www.famitsu.com/news/201802/27152683.html>  
(<https://twitter.com/famitsu/status/968399069189087235> より引用)

我々はこの点に着目し、Twitter において既知の事象が出現した初期の投稿（ツイート）を正例（新固有表現の出現文脈）として収集する。具体的には、i) ある期間に Wikipedia に新規登録された記事のタイトルを新固有表現として収集し、ii) 固有表現ごとに、登録時期の前後の期間でその固有表現を含むツイートを時系列の古い順に収集する。負例については、ii) で集めたツイートの期間の後（一ヶ月後）のツイートを正例と同数収集する。

遠教師あり学習は大量の疑似教師データを生成できる一方で、誤ったラベルのデータが混入してしまう可能性がある。そのため、過去に一定頻度以上言及された記事タイトルを（Wikipedia に登録された時期前後に発生した固有表現ではないとみなして）除いたり、収集する投稿をリツイートされたものに絞って公益性が高い言及に絞ったりなどして誤りの影響を減らす工夫を行う。

続いて収集した疑似教師データを利用して新固有表現を発見するモデルを学習する。提案手法では、新固有表現の検出問題を固有表現抽出のタスクを解く際に標準的に用いられる枠組みである系列ラベリング問題として定式化する。教師あり学習のモデルとしては、系列ラベリング問題で標準的に用いられる手法である条件付き確率場 (CRF) を採用した。モデルで用いる特徴量については、現在位置のトークンとその前後 2 トークンの表層、品詞細分類、文字種、前トークンの出力ラベルを用いる。

最後に提案手法の評価方法について述べる。提案手法は、事前に収集した Twitter 投稿に対して適用して実際に発見した新固有表現の精度によって評価する。まず、具体的に、2012 年 3 月 11 日から 2015 年 12 月 30 日までに日本語版 Wikipedia に登録された記事タイトルで、それ以前の期間の Twitter の投稿に頻度 5 回未満しか現れないものを（その期間に出現した）新固有表現とみなした。次にこれらの新固有表現が出現した投稿を収集し、各固有表現が 1 日に 10 回以上リツイートされる最初の日までの投稿を正例、その後を負例として CRF の学習を行った。

このようにして学習した CRF を 2016 年 6 月 10 日から 16 日までの収集済み投稿に対して適用することで新固有表現を獲得した。得られた新固有表現を獲得回数で整列して頻度上位の 50 件を手で正誤判定し、その累積精度を算出する。

## (2) 実世界テキストを対象とした自然言語処理応用技術

動的に獲得した新固有表現の知識が有用となる自然言語処理応用技術として、i) ソーシャルメディアから大衆の価値観を獲得するシステム（平成 28 年度）、ii) ニューラル対話応答システム（平成 28,29 年度）、iii) ニューラル英日機械翻訳システム（平成 29 年度）の研究開発を行った。これらの手法について簡潔に説明する。

### ソーシャルメディアからの価値観の獲得

まずソーシャルメディアから大衆の価値観を獲得するシステムについて述べる。このシステムでは（様々な）事物を形容詞（観点）で順序付けしたものを価値観とみなす。例えば、ロンドン、パリ、ローマを「安全さ」について順序付けする場合、一例としてロンドン>パリ>ローマのような順序付けを価値観として得ることを目指す。この研究では人々は自身の価値観に基づいて発言を行うと仮定し、ソーシャルメディアテキストに表出する価値観の発露を集めて、大衆の価値観を導出する。

具体的な手法としては、与えられた事物と形容詞に対し、ソーシャルメディアテキストから抽出した順序付けの手がかりを教師あり学習によって適切に組み合わせることで信頼性の高い順序付けを得る手法を開発した。テキスト中に現れた順序付けを示唆する手がかりとしては、各事物と形容詞に関する共起、係り受け、比喩、比較表現を用いる。また、順序付けモデルにはランキング SVM を用いる。

さらに Twitter を対象とし、指定された属性（性別・居住地）のユーザの価値観をオンデマンドで獲得するシステムを開発した。具体的には、事前にプロフィール情報や投稿内容などから Twitter ユーザの属性を得ておき、与えられた属性のユーザが書いた投稿のみを対象として提案手法を適用することで、特定の属性を持つユーザの価値観（順序付け）を獲得できる。

実験では、数十のクエリ（事物の集合と観点のペア）について、クラウドソーシングなどを用いて得た順序付けを教師データとして交差検定でモデルを学習・テストを行い、システムの出力する順序付けと正解の順序付けの相関係数を算出することによって評価する。

### 発話状況を考慮したニューラル対話システム

次に、発話状況を考慮したニューラル雑談対話システムについて述べる。このシステムは、与えられた（ユーザ）発話に対して、適切な応答を生成、または応答候補から選択する。雑談対話においては、発話に対して適切な応答が一意に定まらない場合が多くみられ、統計的なモデルで既存の会話から応答を学習しようとした場合には、適切な学習を行うことが難しいという本質的な問題が存在する。こ

の点を解決することを目的として、発話のトピックやユーザの属性など発話が行われた様々な状況（発話状況）を考慮した応答生成・選択手法を開発した。

提案したシステムでは、ニューラル対話システムにおいて、局所的な発話状況に対して別の応答モデルを用意する手法と、発話状況を疑似的なトークンとして入力発話に加えて応答を生成・選択するという2つの方式を考案し、対話システムにおいて発話状況を考慮する方法論を確立した。

実験では、応答選択タスクを通して提案手法の有効性を評価する。まず、Twitter から抽出したユーザ間の対話（発話一応答ペア）を正解とみなして対話システムの学習を行う。次に、正解の応答を別の会話からランダムに取り出したダミーの応答と混ぜた評価用データセットを用意し、学習で得られた対話システムで正解の応答を選択した正答率によってシステムの性能を評価する。

#### 低コストで学習可能なニューラル機械翻訳

最後に開発したニューラル英日機械翻訳について述べる。現在標準的に用いられるニューラル翻訳システムの多くは、Encoder-Decoder モデルと呼ばれる系列から系列への写像を学習するモデルの亜種である。このモデルでは、パラメタの多さから学習の安定性が低く、パラメタチューニングに多くの時間を割く必要がある。そこで、安定した翻訳性能を示す翻訳システムを低コストで学習するために、事前に教師なしで高速に学習した単語分散表現を用いて Encoder-Decoder モデルの単語埋込み層を初期化する手法を考案した。

実験では、新固有表現が出現しやすい科学技術論文抜粋コーパス (ASPEC コーパス) を用いて翻訳モデルの学習・テストを行い、出力の BLEU スコアによって、その翻訳精度を評価する。

#### 4. 研究成果

##### (1) ソーシャルメディアストリームからの新固有表現の発見手法の確率

研究方法で述べた提案手法により構築した新固有表現発見モデルを代表的なソーシャルメディアストリームである Twitter の投稿（ツイート）ストリームに適用した。提案手法により発見された新固有表現の頻度上位 50 件について、その累積精度を評価したところ、固有表現抽出手法と（Wikipedia に未登録の文字列を除く）フィルタを用いたベースライン手法に比べて、大幅な精度改善を確認することができた。図 1 に、提案手法とベースライン手法によって発見された新固有表現の累積精度のグラフを示す。

なお、ベースライン手法では同名異義の新固有表現（例：映画「ハリー・ポッター」と小説「ハリー・ポッター」）を獲得することはでき

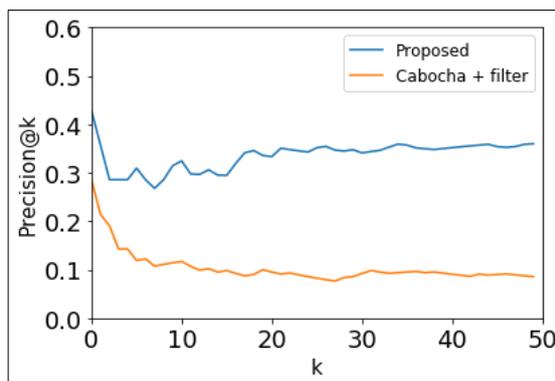


図 1: 獲得された新固有表現の精度@k

きないが、提案手法では獲得可能であることに注意されたい。

以下に実際に獲得できた新固有表現を示す。実際に、固有表現の新規性を周知する投稿から新固有表現を獲得できていることが確認できる。

新固有表現	新固有表現を獲得した投稿
MUFG コイン	三菱東京 UFJ が独自の仮想通貨『MUFG コイン』を発行! 本格的に紙幣を持ち歩かない時代に <u>_URL_</u>
GeForce GTX1070	「GeForce GTX1070」が 6 月 10 日(金)22 時に解禁 一部ショップでは深夜販売を実施 -アキバ総研 <u>HASH URL</u>

表 1: 獲得された新固有表現 (抜粋)

この研究成果については、国内学会・シンポジウム（発表論文 [1,7,13,18]）で公表を行った。

##### (2) 実世界テキストを対象とした言語処理応用の研究

まず、ソーシャルメディアからの価値観の獲得手法については、事前に収集した約 20 億文からなる日本語ブログ記事を知識源として価値観の獲得実験を行い、7 人の被験者から獲得した順序付けを統合した正解の順序付けとの相関係数によってその妥当性を評価した。事物と形容詞の共起に基づくベースライン手法により得られた順序付けの相関係数が 0.274 であったのに対し、提案手法で得られた順序付けの相関係数は 0.441 であり、有意に高い結果であった（学会発表[16]）。

次に、Twitter からの価値観獲得システムについては、C++/Ruby を用いて実装を行い、6 年分 250 億のツイートを知識源として、現実的な時間で動作することを確認した。図 2 は実際に開発したシステムで、映画のアナ雪（アナと雪の女王）、アベジャーズ、パイオハザードを男女別で好きな順番で順序付けした結果である。図から、女性が男性と比べてアナと雪の女王を好む結果が確認できており、シス

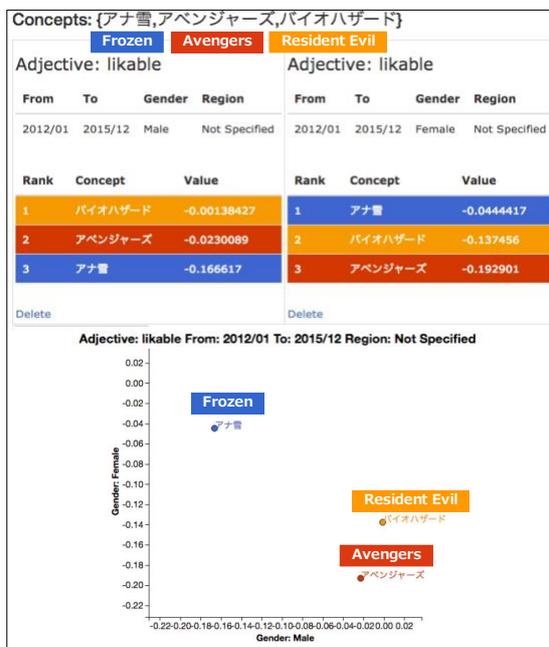


図 2: Twitter からの価値観獲得システム

システムの有用性が確認できる。

この研究成果は国内会議（学会発表[16]）で発表するとともに、国際会議（学会発表[21]）で発表を行い、最終的な成果を論文誌（雑誌論文[1]）に掲載した。また、Twitter からの価値観獲得システムについては、国内シンポジウム（学会発表[19]）および国際学会（学会発表[17]）においてデモ発表を行った。開発した価値観獲得システムについては、オープンソース化し、ウェブサイトで公開している（その他[2]）。

次に、発話状況を考慮したニューラル対話システムについては、Twitter 中の会話を用いた応答生成タスクを通して評価を行い、発話状況を考慮しない場合と比較して有意に応答選択性能が向上することを確認した。

この研究については、途中段階での成果を国内会議（学会発表[23]）で発表した。構築した対話システムをもとに、評価型国際ワークショップ NTCIR-12（学会発表[22]）の Short Text Conversation タスクに参加して 7 チーム中 3 位の成績を得ている。またこの国際ワークショップでは成績上位者として口頭発表も行っている。最終的な研究成果については、国内会議（学会発表[11, 12]）で発表するとともに、国際ワークショップ（学会発表[9, 21]）で発表している。

最後に、低コストで学習可能なニューラル英日機械翻訳システムについては、国内研究会（学会発表[6]）でその成果を発表するとともに、評価型国際ワークショップ The 4th Workshop on Asian Translation (WAT 2017) の科学技術論文英日翻訳サブタスクに参加して、人手の評価で 10 チーム中 6 位の成績を得ている。また、同国際ワークショップでは、参加チームから 3 チームのみが選ばれた口頭発表に選ばれ、成果の発表を行っている（学会発表[3]）。開発した英日機械翻訳システムについ

ては、オープンソース化し、ウェブサイトで公開している（その他[1]）。

## 5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 3 件）

[1] Tatsuya Iwanari, Naoki Yoshinaga, Masashi Toyoda, Masaru Kitsuregawa. **Inducing Writers' Values on Concept Ordering from Microblog**. DBSJ Journal, 16(2), 2018. [査読有](#).

[2] 石渡祥之佑, 鍛冶伸裕, 吉永直樹, 豊田正史, 喜連川優. **文脈語間の対訳関係を用いた単語の意味ベクトルの翻訳**. 人工知能学会論文誌, 31(6), pp. 1-10. 2016. [査読有](#).

[3] 川本貴史, 豊田正史, 吉永直樹. **マイクロブログからの社会的影響力を持つ情報カスケードの検知手法**. 情報処理学会論文誌データベース(TOD), 9(2), pp. 23-33, 2016. [査読有](#).

〔学会発表〕（計 24 件）

[1] 赤崎智, 吉永直樹, 豊田正史. **ソーシャルメディアストリームからの新固有表現の発見**. 第 32 回人工知能学会全国大会. 2018/06/06, 鹿児島市.

[2] 石渡祥之佑, 林佑明, Graham Neubig, 吉永直樹, 豊田正史, 喜連川優. **系列編集モデルに基づく単語ベクトルからの定義文生成**. 言語処理学会第 24 回年次大会. 2018/03/14, 岡山市.

[3] 大原康平, 佐藤翔悦, 吉永直樹, 豊田正史, 喜連川優. **階層型 RNN を用いた対話における応答の対話行為予測**. 言語処理学会第 24 回年次大会. 2018/03/13, 岡山市.

[4] Masato Neishi, Jin Sakuma, Satoshi Tohda, Shonosuke Ishiwatari, Naoki Yoshinaga, Masashi Toyoda. **A Bag of Useful Tricks for Practical Neural Machine Translation: Embedding Layer Initialization and Large Batch Size**. The 4th Workshop on Asian Translation (WAT2017). 2017/11/27, Taipei, Taiwan.

[5] 佐久間仁, 吉永直樹. **単語の表層類似性を用いた多言語単語分散表現の教師なし学習手法**. 第 233 回自然言語処理研究会. 2017/10/25, 宮古島市.

[6] 根石将人, 佐久間仁, 遠田哲史, 石渡祥之佑, 吉永直樹, 豊田正史. **ニューラル機械翻訳における埋め込み層の教師なし事前学習**. 第 233 回自然言語処理研究会. 2017/10/24, 宮古島市.

[7] 赤崎智, 吉永直樹, 豊田正史. **ソーシャルメディアストリームからの多様な新エンティティの検出**. NLP 若手の会第 12 回シンポジウム. 2017/09/04, 那覇市.

[8] 大原康平, 佐藤翔悦, 吉永直樹, 豊田正史, 喜連川優. **対話における応答の対話行為予測**. NLP 若手の会第 12 回シンポジウム. 2017/09/04, 那覇市.

[9] 石渡祥之佑, Jingtao Yao, Shujie Liu, Mu Li,

Ming Zhou, 吉永直樹, 喜連川優, Weijia Jia. ニューラル機械翻訳のための句に基づくデコーダ. NLP 若手の会第 12 回シンポジウム. 2017/09/03, 那覇市.

[10] Shoetsu Sato, Naoki Yoshinaga, Masashi Toyoda and Masaru Kitsuregawa. **Modeling Situations in Neural Chat Bots**. The 55th Annual Meeting of the Association for Computational Linguistics - Student Research Workshop (ACL-SRW 2017). 2017/07/31. Vancouver, Canada. 査読有.

[11] 佐藤翔悦, 吉永直樹, 豊田正史, 喜連川優. 非明示的な発話状況を考慮したニューラル対話モデルの検討. 第 31 回人工知能学会全国大会. 2017/05/23, 名古屋市.

[12] 佐藤翔悦, 吉永直樹, 豊田正史, 喜連川優. 暗黙の発話状況を考慮したニューラル対話モデル. 言語処理学会第 23 回年次大会 (NLP2017). 2017/03/15, つくば市.

[13] 赤崎智, 吉永直樹, 豊田正史. 発生普及過程を捉えた未知エンティティの発見. 第 9 回データ工学と情報マネジメントに関するフォーラム (DEIM2017). 2017/03/07, 高山市.

[14] 小泉実加, 吉永直樹, 豊田正史. マイクロブログにおける画像投稿の文脈に基づくオンデマンド検索手法. 第 9 回データ工学と情報マネジメントに関するフォーラム (DEIM2017). 2017/03/07, 高山市.

[15] 大原康平, 佐藤翔悦, 吉永直樹, 豊田正史, 喜連川優. 不足情報を自律的に問う対話エージェントの実現に向けた聞き返しの必要性検知. 第 9 回データ工学と情報マネジメントに関するフォーラム (DEIM2017). 2017/03/07, 高山市.

[16] Tatsuya Iwanari, Naoki Yoshinaga, Masashi Toyoda, Masaru Kitsuregawa. **Inducing Writers' Values on Concept Ordering from Microblog**. 第 9 回データ工学と情報マネジメントに関するフォーラム (DEIM2017). 2017/03/07, 高山市.

[17] Tatsuya Iwanari, Kohei Ohara, Naoki Yoshinaga, Nobuhiro Kaji, Masashi Toyoda, Masaru Kitsuregawa. **Kotonush: Understanding Concepts Based on Values behind Social Media**. The 26th International Conference on Computational Linguistics (COLING2016) system demonstrations. 2016/12/16, Osaka. 査読有.

[18] 赤崎智, 乾孝司, 吉永直樹, 豊田正史. テキストストリームからの同名異義な未知エンティティの検出. NLP 若手の会 第 11 回シンポジウム. 2016/08/30, 和歌山県西牟婁郡白浜町.

[19] 岩成達哉, 大原康平, 吉永直樹, 鍛冶伸裕, 豊田正史, 喜連川優. **COMAVa: Concept Ordering Method for Acquiring Values**. NLP 若手の会 第 11 回シンポジウム. 2016/08/29, 和歌山県西牟婁郡白浜町.

[20] Masahiko Itoh, Naoki Yoshinaga, Masashi Toyoda. **Word-Clouds in the Sky: Multi-layer Spatio-temporal Event Visualization from a**

**Geo-parsed Microblog Stream**. The 20th International Conference on Information Visualization (iV2016). 2016/07/20, Lisbon, Portugal. 査読有.

[21] Tatsuya Iwanari, Naoki Yoshinaga, Nobuhiro Kaji, Toshiharu Nishina, Masashi Toyoda, Masaru Kitsuregawa. **Ordering Concepts Based on Common Attribute Intensity**. The 25th International Joint Conference on Artificial Intelligence (IJCAI2016). 2016/07/14. 査読有.

[22] Shoetsu Sato, Shonosuke Ishiwatari, Naoki Yoshinaga, Masashi Toyoda and Masaru Kitsuregawa. **UT Dialogue System at NTCIR-12 STC**. NTCIR-12 Conference on Evaluation of Information Access Technologies (NTCIR-12), STC: Short Text Conversation. 2016/06/09, 東京.

[23] 佐藤翔悦, 石渡祥之佑, 吉永直樹, 豊田正史, 喜連川優. 発話状況を意識したオンライン上の対話における応答選択. 第 30 回人工知能学会全国大会. 2016/06/07, 北九州市.

[24] Shonosuke Ishiwatari, Naoki Yoshinaga, Masashi Toyoda, Masaru Kitsuregawa. **Instant Translation Model Adaptation by Translating Unseen Words in Continuous Vector Space**. The 17th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2016). 2016/04/07, Konya, Turkey. 査読有.

[その他]

ソフトウェア

[1] WAT 2017: UT-IIS:

<https://github.com/nem6ishi/wat17>

[2] Kotonush: Understanding Concepts Based on Values behind Social Media.

<http://www.tkl.iis.u-tokyo.ac.jp/~nari/coling-16/>

## 6. 研究組織

### (1) 研究代表者

吉永 直樹 (YOSHINAGA, Naoki)

東京大学・生産技術研究所・准教授

研究者番号： 90773961

### (2) 研究分担者

( )

研究者番号：

### (3) 連携研究者

( )

研究者番号：

### (4) 研究協力者

( )