

倒置法を利用した記事タイトル生成に関する検証*

伊藤巧 琉球大学大学院理工学研究科[†]

赤嶺有平 山田孝治 遠藤聡志 當間愛晃 琉球大学工学部情報工学科[‡]

1 問題背景

World Wide Web (以下 Web) 上に存在する記事 (以下 Web 記事) には、一つの記事に様々な話題が詰まっており、人々が興味を持つ内容が多々含まれている。しかし、Web 記事まとめサイト (Web 記事へのタイトルリンク集) などにある Web 記事へのタイトルリンクは、一つの記事に対して一つしかない。よって、Web 記事内の話題にユーザの興味の対象が含まれていたとしても、一つのタイトルからそれを連想するのが難しい為、見逃してしまう。

例えば、米国時間 2015 年 9 月 9 日にサンフランシスコで開かれた Apple 社の新商品発表会についての Web 記事¹ には、iPad でお絵かきができる Apple Pencil などの発表された新商品についての内容や、発表会に Perfume が出演して話題になったという会場の様子に関する内容など、色々な話題が書かれている。この記事は、Perfume、お絵かきツールなどに興味を持っているユーザにとって、見る価値があると考えられる。しかし、記事へのタイトルリンクが Apple Pencil についてのみ言及していたため、Perfume に興味を持つユーザは自分が興味を持っている内容に気づかず見逃してしまう状況が起こりえる。このような状況を解消したい。

2 研究概要

1 節で説明した問題に対する改善手法の一例として、ユーザに、Web 記事中の話題の中に自身の興味の対象があることを知らせることを提案する。すると、その興味の対象を探し、見つけることができる。よって、自身が興味を持っている内容の記事を見逃すことを防ぎ、問題を解消する事になる。さらに、自身の興味の対象を探す際に、記事全体に目を通す事で、結果的にユーザ

が想定していた対象以外の話題に関して興味を持ってもらう、あるいは知ってもらう事も狙うことができる。

この改善手法を実現するため、本研究では Web 記事中の文書内で話題となっているキーワードの一つを選出し、そのキーワードに関する文書内の一部分を表すタイトルを生成する手法を模索している。将来的には、本研究の手法によって一つの Web 記事から複数のタイトルリンクを生成し、その中からユーザの興味に適した物をユーザ毎に選んで提供するシステムを目指している。

2.1 具体的な研究手法

本研究が過去に行った具体的手法について以下に述べる [1]。まず、Web ページから広告やリンクなどを抜いた Web 記事の文章群のみを抽出した。その文章群から、話題となっている特定のキーワードの一つを選び、そのキーワードをユーザが興味を持つ対象 (以下興味キーワード、今回の場合は “Apple TV” と “Apple Pencil”) と仮定した。そして、興味キーワードを含む文を全て抽出した。抽出された文一つ一つに対し、その内容が明確かつ簡潔になるように筆者がタイトルを生成した。

2.2 文と生成されたタイトルを比較し、文からタイトルを生成する法則を思索

節 2.1 で述べた具体的手法 [1] を踏まえ以下のようにタイトルを生成する法則を思索した。

“～～され、” などの動作の経緯を表す表現は消去する

新たにiOSベースのtvOSが用意され、Apple TV向けにさまざまなアプリが開発・配布できるようになった

興味キーワード “Apple TV” が、述語 “開発・配布できる” を修飾していた場合、興味キーワードに付いている付属語 “向けに” を消して述語の後ろに持ってくる

図 1: タイトル生成法則の検討例

図 1 で示したように、2.1 節で抽出した文章とその抽出文一つ一つから人手で生成したタイトルを比較し、タイトルを生成するための法則を検討した。検討結果を踏まえ、全文に共通して使用できる生成法則全 19 項目に整理した。

*A study of title generation using an inversion

[†]Takumi ITO, Department of Science and Technology Studies, Graduate School of the Ryukyus

[‡]Yuhei AKAMINE, Koji YAMADA, Satoshi ENDO, Naruaki TOMA, Department of Information Engineering, University of the Ryukyus

¹<http://japanese.engadget.com/2015/09/10/ipad-pro-apple-pencil/>

| 法則番号 | 発動条件 処理 |
|------|---|
| 14 | キーワードが主語, または, キーワードに格助詞 (“~では”) が付いている時 キーワードを述語の後ろに移動させる |

表 1: タイトル生成法則 14

表1は全19項目の中で、今回の実験で生成されたタイトルに最も多く使われた生成法則である（以下生成法則14）。よって、この生成法則14からプログラムで実装を試みた。

2.3 生成法則 14 の実装方法



興味キーワードを含むチャンクの係り受け先の一番最後にある動詞の後ろに興味キーワードを持ってくる

図 2: タイトル生成法則 14 の成功例

図2のように、抽出文を係り受け解析器 cabocha で解析し、興味キーワードの掛かり先を特定、そして掛かり先の一番最後の述語（自立語）を判定し、その後ろに興味キーワードを持ってくる処理を行った所、生成法則14の“キーワードを述語の後ろに移動させる”機能を実装することができた。

3 実験結果及び考察

2.3節で作成したプログラムに抽出文（全30文、その内、名詞の羅列だったのでタイトル生成ができなかった文が3文）を掛け、生成法則14を再現してみた。実行結果（全27文）の内、意図していた通りの動作（図2のような動作）を行っているのが10文、意図しない動作をしてしまう文が17文あることが分かった。

3.1 意図しない動作に関する考察

3節で述べた“意図しない動作を行う17文”に対して考察を行った所、大まかに2つのタイプに分類されることが分かった。以下にそのタイプを例と共に示す。



図 3: 興味キーワードの直前の単語への処置

図3で示すように、興味キーワードの前に形容詞や動詞（今回の場合は“新しい”）がついているタイプの文では、意図している動作として、興味キーワードと一緒に移動して欲しい。だが、2.3節のプログラムでは興味キーワードしか動かす対象にしていない。よって、プログラムを書き換え、興味キーワードの前の単語も解析し、興味キーワードと同じように動かすべきかを判定するようにならなければならない。

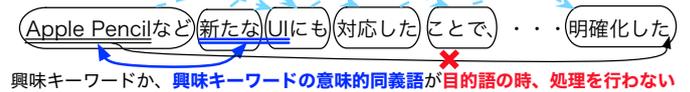


図 4: 興味キーワードが目的語だった場合

図4で示すように、興味キーワードが目的語、または目的語の代名詞（今回の場合は“新たな UI”）と同じ意味を示したタイプの文では、意図していた動作として処理をしないのが正解だが、2.3節のプログラムでは、文中での興味キーワードの働きを考慮していないため、全ての文に対して同じ動作を行ってしまう。よって、プログラムを書き換え、文中での興味キーワードの働きを解析し、それによって、動作を行うかどうかを判定する必要があると考えられる。

4 今後の予定

3節で述べたような意図しない動作を修正し、現在試している全30文の抽出文に対する生成法則14のプログラムを完成させる。その後、残りの18項目の生成法則に対しても、同じように生成プログラムを作り、全30文に対して、19項目の生成プログラムを作成する。その後、サンプル数を増やして実験を行い、項目の修正、あるいは新しい項目の追加などを検討していきたいと考えている。

また、2節で記されている“抽出文の内容が明確かつ簡潔になるようなタイトル”の評価は、著者の主観に頼っているので、“内容が明確かつ簡潔になっている”かどうかを評価する方法を模索したいと考えている。

参考文献

[1] 伊藤巧, 當間愛晃 (2016) 「ユーザの興味に応じた記事タイトル自動生成に関する基礎研究」, 情報処理学会 第78回全国大会講演論文集, (2R-06, 分冊2):573-574 2016(Mar.)