2R-06

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*

伊藤巧 當間愛晃 琉球大学工学部情報工学科†

1 問題背景および研究概要

人にとって、世界に存在するあらゆる物事に対し興味を持つ事は、好きな物を見つけ、それが趣味などの日々の生きがいや将来就きたい仕事になる。また、自分にとって思いもしなかった考え方や物の見方を見つけ、日々の生き方に変化を与えたり、仕事に対するモチベーションを上げるなど、あらゆる面で人生を豊かにする要因になっている。筆者はこれらの根源である興味の幅を広げることに興味がある。

現在世界中で普及されている World Wide Web（以下 Web）には、限りなく多くの情報が公開され、今なお増加中である。この 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用意されていない。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、素通りしてしまい、結果として見逃してしまう。

例えば、米国時間 2015 年 9 月 9 日にサンフランシスコで Apple 社の新商品発表会が開かれた。それについての Web 記事[1] には、iPhone 6s の機能性やデザイン、iPad でお絵かきができる Apple Pencil などの発表会で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ているという発表会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、見る価値のある物であると考えられる。しかし、記事へのタイトルリンクが Apple Pencil の事のみについて言及されている物であったため、Perfume に興味を持つユーザは自分が興味を持っている Perfume の内容に気づかず素通りする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。

改善手法の一例として、ユーザに、タイトルリンク先の Web 記事に含まれている話題の中にユーザの興味の対象の存在を知らせることを想定している。すると、ユーザは自身の興味の対象を探し、見つけることができる。よって、ユーザの興味に関する内容を含む記事を見逃すことを防ぎ、現象を解消する事になる。さらに、自身の興味の対象を探す事により、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興味の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っている。これを実現するため、本研究では Web 記事内の文書を解析し、ユーザの興味を示すキーワード群に関して書かれている部分を抽出する。次に抽出した文書の内容を表現する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、その中からユーザが持つ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 先行研究

先行研究に、慶留間ら [1] の研究がある。この研究では、Hierarchical Dirichlet Process（HDP）モデルとギブスサンプリングを組み合わせたアルゴリズムを使い、Web 記事中の文書をトピック別に分割する。そして、文書中の単語を組み合わせて生成した語句から、N-gram 言語モデルによって、それぞれのトピックに適したラベルを決定しようとしている。この生成されたラベルは文章中の単語を組み合わせから作られているという点で、本研究のタイトル生成に似ている。しかし一方で、この研究は、複数の似通った Web 記事群から、似通った話題群に対するラベル生成を検討しているので、一つの Web 記事中の内容に書かれている興味キーワードの説明を促すタイトルを生成しようとする本研究とは、全く異なる。よって、本研究で、この研究のラベル生成技術がそのまま使えるとは考えにくいという可能性があるため検討の余地がある。

3 タイトル生成手法を検討する手順

本研究は、Web 記事中の話題に関するキーワードの中から一つ選び、そのキーワードをユーザが興味を持

*A basic study to generate a title for user's interest

†Takumi ITO, Naruaki TOMA, Department of Information Engineering, University of the Ryukyus

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

Copyright ©2016 Information Processing Society of Japan.
All Rights Reserved.

つ対象と仮定する。そして、そのキーワードを使い、よりユーザが興味を持つためのタイトルを生成する手法を検討する。

よりユーザの興味を持つためのタイトルを明確にするため、まずは一つのキーワードから複数のタイトルを人手で生成する。これは、同じキーワードから作られたタイトルでも、そのキーワードをどのような着眼点で見るか、または、どこを強調するかによって、タイトルの内容は各々変わってくるからである。よって、様々な内容のタイトルを観察し、それぞれのタイトルにどのようなユーザが、より興味を持つかを考察した。

3.1 Web 記事から特定のキーワードを含む文章を抽出

小野裕作氏が作成したプログラム [2] を使って、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を一つ選び、そのキーワードが使われている文章を全て抽出した。

3.2 抽出された文章一つ一つに対し、タイトルを作成

・Game Centerでユーザー認証を行うことで、iPhone/iPadでプレイ中のゲームの続きを、Apple TVで行うといったことも可能になるという
→iPhone/iPadでプレイ中のゲームの続きをApple TVで行うことが可能

・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
→さまざまなアプリが開発・配布できるようになったApple TV

図 1: 抽出文と生成したタイトル例

上図で示したように、抽出された文章一つ一つに対し、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。

3.3 文章と生成されたタイトルを比較し、文章からタイトルを生成する法則を思索

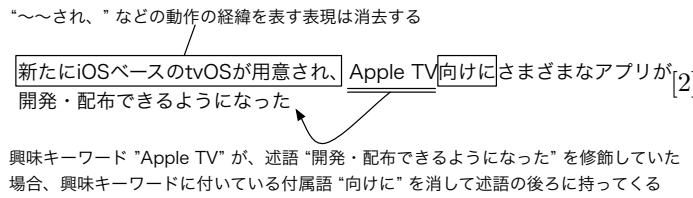

図 2: タイトル生成法則の検討例

上図で示したように、抽出文から、3.2 節で生成したタイトルを生成するための法則を検討した。

4 分析結果および考察

今回抽出した文章、全 25 文について、3.1 節〜3.3 節を行い、全文章に共通して使用できるタイトル生成処理（全 19 項目）を考えた。抽出された文章を見てみると、文章一つ一つに書かれている内容は興味キーワード（今回の場合は “ Apple TV ”）の事について書かれているが、その着眼点は様々である。抽出文の着眼点を残しつつ、興味キーワードを含んだタイトルを生成すれば、そのタイトルに惹かれるユーザの種類が変わるはずだと筆者は考えている。

図 1 を例にとると、「iPhone/iPad でプレイ中のゲームの続きを Apple TV で行うことが可能」というタイトルには、iPhone や iPad、またはスマートフォンでソーシャルゲームに勤しむ人が、「さまざまなアプリが開発・配布できるようになった Apple TV」というタイトルには、アプリ開発者やコンサルティング会社に勤める人が、それぞれのタイトルに惹かれると考えられる。

5 今後の予定

今後は、今回の研究結果により考えだされたタイトル生成処理（全 19 項目）を実装する。その後、文章サンプル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、3.2 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。その評価とタイトル生成処理が実装されたプログラムによる実験結果を考慮して、タイトル生成処理の項目を検討しようと考えている。

参考文献

[1] 慶留間ら（2015）「可変長 N-gram に基づいたトピックへのラベル選択の検証」, 情報処理学会 第 77 回全国大会講演論文集, 2015 年 1 号, pp.217-219.

[2] 小野裕作（2011）「extractcontent.rb の python 版」, <https://github.com/yono/python-extractcontent> (参照 2015-11-20)

Copyright ©2016 Information Processing Society of Japan.
All Rights Reserved.