

リード文と重要語句に基づく記事見出しの忠実度評価

片倉 多智^b 木村 優平^b 戎 淳^b 吉原 圭祐^b 早坂 絵央^b 西村 百之輔^b 延澤 志保*^b東京都市大学知識工学部

*東京都市大学情報工学部

1 研究背景

読者の興味を引くため誇大見出しを付けた記事(クリックベイト)の問題について, グノシー等の大手 web サイトでも排除の取り組みが行われている [1]. 本研究では, クリックベイトによって読者の情報収集を妨害されないよう, 記事見出しの適切さを自動評価する手法を提案する.

2 記事見出しの忠実度評価手法

2.1 提案手法の概要

要約システムによって生成される要約が満たすべき原則として, (1) 要約から情報を読み取れること, (2) 元の文章が要約を含意すること, (3) 要約から読み取れる情報が重要と思われること, の3点から成る3原則 [2] が提案されている. 本研究の対象とする大手ニュースサイトの見出しが原則 (1) を満たさないことは考えにくく, 本研究では原則 (1) については満たされているものと仮定する. 提案手法では入力として評価対象とする見出し, 対応する記事本文, 比較記事群の3種類をとり, 出力は本文内容への忠実度に基づく評価点とする (図1). 評価点

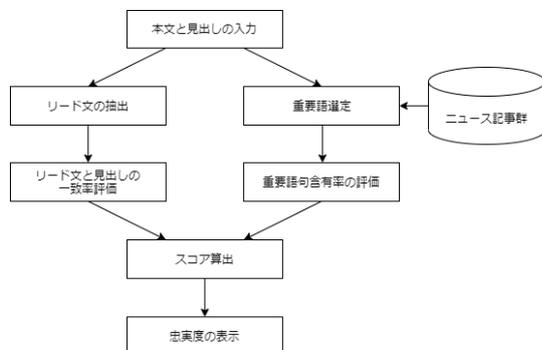


図1: 見出し自動評価の処理の流れ

の算出では, 原則 (2) の評価としてリード文一致率, 原則 (3) の評価として重要語句含有率を用いる.

2.2 リード文一致率による内容反映度評価

記事の中には, 1,000文字以上の本文を20字程度の見出しに要約しているものも存在するため, 本文をそのまま一致率の算出対象として扱うことは適切ではない. 本研究では評価尺度の1つとして記事のリード文と見出しの一致率を用いる. 図2にある記事のリード文の例を挙げる. 石垣らは, 質問文の要約とリード文との比較評価

記事1: リード文

アメリカのトランプ大統領は11日バイデン氏が大統領選での勝利を確実にしてから初めての公務に臨みましたがこの日も発言はなく沈黙を続けています

図2: 記事リード文の例

を行うことで, リード文の参照要約としての有用性を確認した [3]. 見出しも要約の1つであり, リード文との比較による妥当性の評価は可能と考えられる.

図2のリード文に対して, 表1の4種類の見出しを考える. 表1の見出し a は図2の記事の実際の見出し, b

表1: リード文と見出しの語の重複率算出例

見出し	リード文重複率
a トランプ氏6日ぶり公の場沈黙続ける	6/10= 0.600
b アメリカ現大統領バイデン当確に沈黙	5/7= 0.714
c 選確実を初めてから臨み	6/6= 1.000
d 沈黙を続けています	6/6= 1.000

は重要語句を考慮した見出し, c はリード文の形態素をランダムに抜粋したもの, d はリード文の一部抜粋である. 表1ではリード文に含まれる形態素を下線で示してある. リード文と見出しとの重複率では, 見出し a よりも見出し c と見出し d の方が数値が高い (表1). これは見出し c と見出し d がリード文中の形態素を並べてあるため, この2種類が見出し a に比べて適切なわけではないことは明らかである. このように, リード文との含有語の重複率では見出しの妥当性を測ることはできない.

本研究では, 見出し c のようなランダムな語の並びを適切に評価するため, リード文一致率の算出にあたって文脈を考慮する手法を提案する. 2つのテキストについて順序を考慮して一致率をとる代表的な方法である ROUGE-N は, システムの出力した要約を参照要約と比較する評価尺度で, 参照要約とシステム要約の間で一致する n -gram の割合を計算する [4]. Lin らは, n を1から4まで変化させてマニュアルでの評価結果との相関を調べた結果 $n=1$ および $n=2$ が最も高い相関であったと報告している [4]. 石垣らは $n=2$ で成果を出しており [3], 提案手法でも文脈を考慮するため, リード文との一致率として ROUGE-1,2 とそれらの F 値を利用する.

2.3 重要語句含有率評価による要約重要度評価

短い要約文では語句選択が精度向上に影響する. 例えば, 「米国」や「大統領選挙」という語句が含まれていなくても, 「バイデン候補」という語句を含む2020年の記事であればアメリカ大統領選のニュースと推察できる (表1). このような語句を提案手法では重要語句と考え, これをより多く含む見出しが本文に対して忠実度が高いも

Fidelity Evaluation of News Headlines Based on Lead Sentences and Keywords.

Taichi Katakura^b, Yuhei Kimura^b, Bo Rong (戎 ポツ), Keisuke Yoshihara^b, Kaio Hayasaka^b, Momonosuke Nishimura^b, and Shiho Hoshi Nobesawa*.

^b Faculty of Knowledge Engineering, Tokyo City University

* Faculty of Information Technology, Tokyo City University

のとして評価する。本稿では重要語句を形態素単位で扱うが、これを語句単位に拡張することで複合語や言い回しの反映が期待できる。

提案手法では、重要語句の推定手法として、Okapi BM25[5]を基に、見出しに出現する各形態素 w について対応する記事本文 d での重要度スコア (式 (1)) を提案する。ここで D は記事群、 $avgdl$ はすべての記事の平均の形態素数、 k と b はチューニング用の変数で、ここでは $k = 2.0$ 、 $b = 0.75$ とする。TF は形態素 w の文書内頻度、IDF は形態素 w の逆文書頻度を示す。

$$\text{sig}(w, d) = \text{IDF}(D, w) * \frac{\text{TF}(w, d) \times (k + 1)}{\text{TF}(w, d) + k \times (1 - b + b \times \frac{|D|}{avgdl})} \quad (1)$$

表1の4種類の見出しに含まれる重要語句の割合を表2に示す。下線は重要度スコアが一定の値以上の形態素であり、これを重要語句とする。表2に示すとおり、図2

表2: 見出しの重要語抽出例

見出し	重要語句含有率
a トランプ氏6日ぶり公の場沈黙続ける	6/10= 0.600
b アメリカ現大統領バイデン当確に沈黙	3/ 7= 0.429
c 選 確実を初めてから臨み	3/ 6= 0.500
d 沈黙を続けています	1/ 6= 0.167

の記事の実際の見出しである見出し a が最も高い重要語句含有率を示す。2.2 節のリード文一致率評価では高得点を得た見出し d は重要語句の評価では値が低く、見出し d は重要語句の含有率の点で問題があることがわかる。

表1, 表2のうち、リード文から作成されていない見出し b は、リード文一致率、重要語句含有率ともに見出し a に比べて低くなるが、見出しとして不適切な見出し c、見出し d と比較すると重要語句含有率が高い。重要語句は記事内で明らかに多く出現するとは限らず、本研究では重要語句の推定に比較記事群を用いている。リード文との関連の薄い見出し b を適切に評価できたのもこのためである。記事では、「アメリカ」のような広いカテゴリから「大統領選挙後のトランプ大統領の沈黙」のような関連記事内での特徴まで、重要語句にも粒度がある。見出し a に対して比較記事群を基に各形態素の重要度を算出した結果を図3に示す。ここでは比較記事群として、対

関連記事	トランプ	氏	6	日	ぶり	公	の	場	沈黙	続ける
	0.005	0.002	0.000	0.002	0.029	0.025	0.000	0.023	0.029	0.015
ランダム	トランプ	氏	6	日	ぶり	公	の	場	沈黙	続ける
	0.066	0.041	0.000	0.01	0.016	0.022	0.000	0.022	0.025	0.019

図3: 比較記事群を考慮した重要語の抽出例

象記事の関連記事群と、ジャンルをランダムに選んだ前1週間以内の記事100件との2種類を用いた。図3に示すとおり、関連記事群を基に重要語句を推定した場合、形態素「沈黙」が特徴的であるとして高いスコアが出た半面、形態素「トランプ」はほぼすべての記事に登場するため逆文書頻度が低くなりスコアが小さくなり、後半部分が重要と推定されたことがわかる。それに対して、ランダム記事群を基に重要語句を推定した場合には、記事の内容を明確に示す形態素「トランプ」が最も高いスコア

を出しており、「沈黙」等の述語がそれに続いている (図3)。本研究では広いカテゴリでの重要語句の推定について評価するため、重要語句含有率の算出にランダム記事群を用いる。今後の課題として、広いカテゴリでの重要語句と関連記事内での差別化のための重要語句との適切な組み合わせ手法の確立が望まれる。

2.4 リード文と重要語句を考慮した忠実度算出

提案手法では、記事要約としての適切さをリード文との一致率 (2.2 節) から、記事内容の表現としての妥当性を重要語句の含有率 (2.3 節) から推定し、これらを考慮して記事本文に対する見出しの忠実度評価を行う。忠実度を算出する数式を式 (2) で定義する。

$$\text{fid}(h, d) = (F(h, ls) + \text{ROUGE}_2(h, d)) * \frac{|\text{sig}|}{dl} \quad (2)$$

F はリード文と参照要約の ROUGE の F 値である。 h は対象の見出しであり、 ls はリード文である。 $|\text{sig}|$ は重要語と判定された語句の数、 dl は文書に含まれる形態素の総数である。

表3にリード文一致率と重要語句含有率に基づく忠実度評価の結果を示す。表3から本来の見出しである a の

表3: 記事見出し忠実度評価結果

見出し	ROUGE-2		重要語句	忠実度
	リード文	本文		
a	0.280	0.333	0.600	0.329
b	0.275	0.000	0.429	0.118
c	0.249	0.000	0.500	0.125
d	0.327	1.000	0.167	0.226

忠実度が最も高く算出されており、内容を反映しているものの要点を捉えられていない d がそれに続いている、意味が通じない c や本文の内容を反映していない b はさらに低いスコアとなっており、本研究の目的を達成しているスコアを算出できていることがわかる。

3 まとめ

インターネット上の記事の見出しには本文の内容と乖離した不適切なものも存在する。本研究では、見出しの忠実度を評価する手法として、リード文との一致率と、本文とその関連記事から抽出した重要語句の含有率に着目した2段階の処理による評価方法を提案した。

参考文献

- [1] 関喜史, 吉田光男, “クリックベイトなニュース記事の特定に向けたネガティブフィードバック分析,” DEIM Forum, no.A7-2, 2019.
- [2] 西川仁, “自動要約における誤り分析の枠組み,” 自然言語処理, vol.23, no.1, pp.3-36, 2016.
- [3] 石垣達也, 高村大也, 奥村学, “複数文質問を対象とした抽出型および生成型要約,” 自然言語処理, vol.26, no.1, pp.37-58, 2019.
- [4] Chin-Yew Lin, Eduard Hovy, “Automatic Evaluation of Summaries Using N-gram Co-Occurrence Statistics,” In Proc. of the Human Technology Conference (HLT-NAACL), pp.71-78, 2003.
- [5] Stephen E. Robertson, Zaragoza Hugo, “The Probabilistic Relevance Framework: BM25 and Beyond,” NOW Publishers Inc., 2009.