

## D-39 ユーザの選好と多様性を考慮した記事選択手法の検討

日高 航太朗, 嶋田 和孝  
九州工業大学大学院 情報工学府

## 1 はじめに

近年, ニュースをスマートフォンなどを通じて閲覧する機会が増加している. 既存のニュース配信サービスでは, ユーザはニュース記事の見出し一覧から興味のある記事へアクセスし情報を得る. しかし, ニュース記事の見出し一覧から興味のある記事を探すことはユーザにとって手間のかかる作業である. そのため, 本研究では図 1 に示すような 1 つの画面で完結する紙面要約を行い, ユーザがより簡単に情報を得ることができるようなレイアウトを作成することを目的としている. 1 つの画面上で完結するレイアウトの作成にあたって, 以下の 2 点を本稿での課題とする.

- 内容に偏りの無い網羅性の高い記事選択
- ユーザの選好を考慮した記事選択

本研究では, この 2 つを考慮した記事選択モデルを作成し, 比較評価を行う.

## 2 データセット

本研究では, 西日本新聞の作成した新聞記事データを利用する. 本データには記事見出し 1 つにつき 1 つのカテゴリタグが付与されている. カテゴリタグには, 福岡県, 大分県など地域に関するタグの他, スポーツ, 政治, などの記事内容のジャンルに関するタグ, 速報, コラムなどがある.

## 3 提案手法

網羅性の高い記事選択を行うために, 本研究では MMR を利用する [1]. MMR では, 初めに記事選択を行う対象の記事集合  $D = \{D_1, D_2, \dots, D_m\}$  の中から記事の重要度  $Imp$  が最も高い記事を選択する. その後の記事選択では, まだ選択されていない記事に対して, 既に選択された記事集合  $S = \{S_1, S_2, \dots, S_n\}$  との類似度  $Sim$  と記事の重要度を用いて新たなスコアを算出し, 最もスコアの高い記事を選択する. このスコアを用いることにより, 記事の重要度だけでなく類似性を考慮した網羅性の高い記事選択が可能になる. スコア算出の際にはパラメータ  $\lambda$  を用いて重要度と類似度のどちらに重きを置いて記事選択を行うか調整することが可能である. MMR の具体的な計算式は (1) 式ようになる.

$$SCORE_{MMR} = \arg \max_{D_i \in D \setminus S} [\lambda Imp(D_i) - (1-\lambda) \max_{S_j \in S} Sim(D_i, S_j)] \quad (1)$$

ユーザの選好を考慮した記事選択を行うためには, ユーザの閲覧履歴が必要となる. 今回は, データセットに存在するカテゴリタグを 1 つ選ぶ. 選択したカテゴリタグが付与されている記事選択対象の日付の前日記事の中から記事を 1 件ランダムに選択し, これを擬似的なユーザの閲覧履歴として利用する. その後, 選択された擬似的なユーザの閲覧履歴  $D_c$  との類似度を算出し, MMR の計算式に加算する. これにより, ユーザが過去に閲覧した記事と類似した内容の記事を優



図 1: 本研究のイメージ図.

表 1: 提案手法による選択結果.

選択順位	提案手法
1	サイが 3 年ぶりにやって来る年明け目処, 福岡動物園が 2 頭公開
2	19 年は... ラグビー W 杯の年でした日田市「15 の出来事」
3	降車を助けてくれた外国人
4	東海大福岡高, 女子サッカー部 3 年連続全国へ初 V 目指し練習に熱
5	「政治とカネ」疑惑の雲隠れ 3 議員ボーナス満額支給に批判も
6	腰痛緩和装置で起業佐賀大生らが共同開発
7	無声映画と活弁, 体験を長崎市で定期鑑賞会, 12 日第 1 弾
8	宮崎県三股町で建物火災の情報
9	教会コンサート始まる新上五島町, 15 日まで 6ヶ所
10	脳梗塞, まひ克服の大衆演劇役者「待つ人のために舞台へ」

先的に選択することが可能になる. 提案手法の具体的な計算式は (2) 式ようになる.

$$SCORE_{suggest} = \arg \max_{D_i \in D \setminus S} [SCORE_{MMR} + sim(D_i, D_c)] \quad (2)$$

## 4 実行例

西日本新聞の発行したニュース記事データを用いて, 対象の日付に発行された記事集合から記事選択を行う. 2019 年 12 月 10 日の記事を対象に, 提案手法を利用して記事選択を行った結果を表 1 に示す. 今回, 提案手法においてユーザの閲覧履歴として選択された記事はカテゴリタグ「福岡」が付与されているプロサッカーチームに関する記事である.

表 1 の結果から, 4 位に福岡県のサッカーに関連する記事や, 2 位にスポーツに関連した記事が選択されている. このことから, 擬似的に仮定したユーザの選好を考慮して記事選択が行えていることが確認できた. また, 3 位にはコラム, 5 位には政治に関する記事, 6 位と 7 位には佐賀や長崎の地域に関する記事が選択されている. これらの結果から, MMR によってユーザの選好に合わせて福岡の記事ばかりを選択するのではなく, 多様な内容の記事選択を行えていることが確認できた.

## 5 おわりに

本研究では, 記事内容の網羅性とユーザの選好を考慮した記事選択手法の提案を行った. 今後は記事選択手法の改善や文字列長を考慮した記事本文及び記事見出しの要約生成に取り組んでいく予定である.

## 参考文献

- [1] Jaime Carbonell, Jade Goldstein. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st annual international ACM SIGIR*, pp. 335–336, 1998.