

報道の多様性を考慮したニュース閲覧システムの提案

切通 恵介[†] 許 洽[†] 馬 強[†]

京都大学大学院情報学研究科[†]

1. はじめに

ニュースは報道元によって内容の差異が大きく、内容も多様性に富んでいる。そのため、ユーザはニュース報道を比較しつつ出来事を理解することが重要である。ニュースの理解を支援するため、多様性に着目したニュースの分析手法やシステムは多く提案されている[1], [2]。

我々は、ニュース記事に対する人物・組織・地域などの Named Entity に着目したニュースの差異分析手法について研究開発を行っている。先行研究[2]では、我々はニュース記事間の差を4つの尺度（関連度、発散度、擁護度の差、詳細の差）で定義し、その定量分析と比較記事のランキング手法を提案している。本研究ではこれらの手法を拡張し、ユーザの閲覧履歴を考慮して、現在閲覧中の記事に対する差分の最も大きい記事を、関連度と組み合わせた3つの視点（発散度、擁護度の差、詳細の差）から推薦する。そして、提案システムをニュースアプリとして開発する。さらに、先行研究[2]の有効性を大規模な実験によって示す予定である。

2. 提案システム

我々の提案するニュース閲覧システムは、ニュースのより多様な視点を効率的にユーザに与えることを目的としている。

パソコンでのニュース閲覧は多く時間を割けるため、興味のある記事に対しては検索やリンクを辿るといった方法で読み比べがしやすい。一方、スマートフォンでのニュース閲覧は移動中のような時間の無い場面で行うことが多く、能動的に関連記事を検索して閲覧を行うことは少ない。また、スマートニュースなどのニュースアプリがよく利用されるが、1トピック1記事しか提示されず、多様なニュース報道に効率よくアクセスできない場合がある。

そこで、本研究では様々な関連記事を効率よくアクセスできる、スマートフォン上の多様性指向のニュース閲覧システムを開発する。

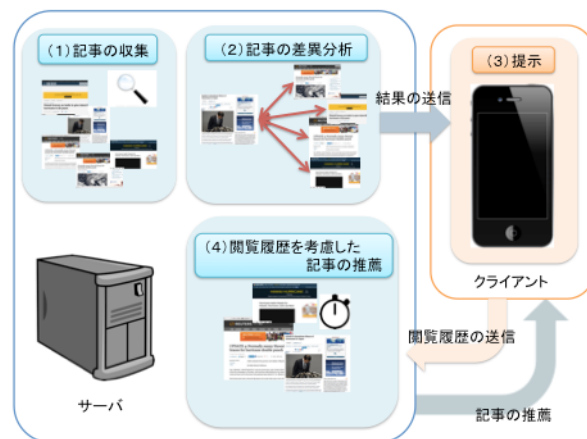


図 1 システム構成

図1に示されているように、提案システムのシステム構成は(1)記事の収集、(2)記事の差異分析、(3)提示と(4)閲覧履歴を考慮した記事の推薦の4つの部分から構成される。

2.1 記事の収集

提案システムにおいては、英語記事を対象とし、記事を Google News (<https://news.google.co.jp>) から収集する。各トップ記事と、その記事の Realtime Coverage から関連記事を収集する。これらの記事は提案システムの入力とする。

2.2 記事の差異分析

先行研究[2]の手法を拡張し、トップ記事とその関連記事の差異を以下の三つの尺度（発散度・擁護度の差・詳細の差）を用いて分析する。

発散度は記事間での視点違いの程度を表す尺度、擁護度の差は Named Entity に対する記事間の感情の差を示す。また、詳細の差は Named Entity に対する詳しさの記事間での差を示す。先行研究[2]で示された結果を用いて、記事間の関連性を示す関連度と組み合わせた発散度・擁護度の差、単独で計算した詳細の差の3つの尺度から記事間の差を計算する。

2.3 提示

提案システムはトップ記事表示画面と詳細記事表示画面の2つの画面から構成される。インターフェースの例を図2に表す。

トップ記事表示画面では Google News から収集したトップ記事をカテゴリ毎にまとめ、サムネイルと見出し、概要を列挙する。各記事を選択

することで記事の詳細表示画面へ遷移する。

詳細表示画面では選択された記事のリンク先を表示する。画面内の「異議あり」「ワイドアングル」「深読み」の3つのボタンが、それぞれ擁護度の差、発散度、詳細の差に対応している。ボタンを選択すると現在閲覧中の元記事の関連記事集合のうち、対応する尺度の値が最も大きい記事を表示する。

2.4 閲覧履歴を考慮した記事の推薦

トップ記事に対する差分のみから各尺度における記事を表示する場合、遷移した記事から得られる情報の量が小さくなる可能性がある。例えば、「異議あり」と「ワイドアングル」が全く同じ記事であった場合、ユーザは各ボタンから遷移をしてもニュースの多様な視点を得られない。本研究ではこの問題に対してユーザの閲覧履歴を考慮して各尺度の値を更新することで関連記事の再ランキングを行う。

記事から得られる多様性を向上させるには、閲覧した記事それぞれから見た尺度の値が最も大きな記事を得られればよい。トップ記事 o の関連記事集合 A に対して、ユーザの閲覧履歴 H_i ($H_i \subset A$) を考える。記事 b ($b \in \{A - H_i\}$) の各尺度の値 $up_def(b, H_i)$ を以下の式で更新する。

$$up_def(b, H_i) = \frac{1}{|H_i|} \sum_{h \in H_i} def(b, h_i) \quad (1)$$

ただし、 $def(b, h_i)$ は記事 h_i から見た記事 b の各尺度の値である。また、 $|H_i|$ は閲覧した記事の個数である。この方法では記事 h_{i+1} を閲覧する度に $def(b, h_{i+1})$ が計算されるため、記事数が多い場合には特に計算量が大きくなる。この問題に関して、記事集合 $A - H_i$ に対して計算する記事数を絞ることで対処を行う。直感的に、トップ記事との計算においてランキング下位にある記事は、履歴によるランキングの更新後も下位にあると考えられる。この仮説が成り立つならば、ランキング下位の記事の計算を省くことが出来る。この仮説を確かめるために予備実験を行った。

Google News から収集した同じニュースイベントの記事0から15の計16記事に対して、履



図2 トップ記事表示画面と詳細表示画面

歴を考慮した発散度の計算を行った。閲覧記事の選択をランダムに6回行い、選択する度に(1)式に基づきランキングを更新した。各選択により更新されたランキング上位8件を表1に示す。

表1より、履歴の増加によるランキング上位での大きな変化は見られなかった。従って、ランキングに関する仮説が限定的に成り立つことが推測される。

3. 実験計画

我々はクラウドソーシングによる差異分析の評価と提案システムのユーザビリティ評価を行う予定である。

差異分析の評価に関して、ニュース記事はGoogle News US版のTop Storiesのうち、2014年6月17日から6月24日までの40ニュースイベントに対してRealtime Coverageから収集した記事を用いる。これらから元記事を1記事、関連記事を平均15記事選択する。被験者はアメリカ人に限定し、ニュース比較のタスクを提示する。被験者は、まず、元記事と関連記事を読む。次に、元記事と関連記事と比較し、先行研究[2]で定義された4つの指標の強さを5段階で評価する。このタスクを1ニュースイベントに対して10名ずつでの評価を行う。各タスクの所用時間は1分辺り250~300単語として記事の単語数から設定する。

ユーザビリティ評価は、多様性を考慮したニュースの理解を主眼に調査する予定である。

4. まとめと今後の課題

我々は、本稿で先行研究[2]を用いたニュース閲覧システムと、閲覧履歴を考慮した再ランキングの手法を提案した。今後は計画中の2つの評価と、ランキングに関する仮説を数学的な証明を行い、システムに反映させたい。

参考文献

- [1] S. Park, S. Kang, S. C. J. Song.: Newscube: delivering multiple aspects of news to mitigate media bias, CHI 2009, pp. 443-452 (2009).
- [2] K. Kiritoshi and Q. Ma: "Named Entity Oriented Related News Ranking", DEXA 2014, LNCS 8645, p.83-98

表1 記事の遷移によるランキングの変化

順位	初期値	1回目	2回目	3回目	4回目	5回目	6回目
1	記事3	記事3	記事3	記事3	記事3	記事3	記事3
2	記事10	記事1	記事1	記事1	記事1	記事1	記事1
3	記事1	記事5	記事2	記事2	記事2	記事4	記事4
4	記事5	記事2	記事4	記事4	記事4	記事12	記事12
5	記事4	記事4	記事12	記事12	記事12	記事11	記事11
6	記事2	記事12	記事15	記事15	記事11	記事7	記事7
7	記事12	記事15	記事11	記事11	記事7	記事8	記事8
8	記事15	記事11	記事13	記事7	記事8	記事6	記事14