

文章に対する興味の類推を目的とした視線情報の活用

小川 拓也† 遠藤 慶一† 小林 真也†

†愛媛大学大学院 理工学研究科

1 はじめに

インターネット上には膨大な量のニュース情報があふれているが、そのすべてがユーザにとって興味のあるものとは限らない。ユーザがニュースを読むとき、ユーザは膨大なニュース情報の中から興味のある情報を探し出す必要がある。情報量の増加に伴い、ユーザによる情報の選別が困難になることを情報過多という。情報過多の問題を解決することを目的として、個人向け情報配信システム PINOT (ピノ: Personalized INformation On Television screen) が開発されている [1][2][3]。PINOT は、ユーザの情報選択操作からユーザの興味を学習し、学習した興味を考慮した情報提供を行うことで、ユーザによる情報選別を不要とし、情報過多の問題を解決していた。しかし、興味の有無の判定を記事単位で行っているため、単語に対する興味の類推としては、信頼性が低い。そのため、ユーザプロフィールの作成には、膨大な量の記事選別が必要となり、ユーザの負担となっていた。そこで、より詳細な興味の度合いを取得する手法を確立することが求められる。

2 視線を考慮した興味学習手法

より詳細な興味の度合いを取得するために、ユーザの視線を考慮した興味学習アルゴリズムを提案する [4]。ユーザが文中のどの単語に注目していたかによって、興味の学習の記事単位から単語単位へと、より小さな単位で行うことが可能になり、関心をより正しく反映した学習ができると考えられる。また、注視のし方に興味の強さに応じた特徴が現れるとすれば、その傾向を明らかにすることで、ユーザの意識的な操作に依存することなく興味の学習が可能になると考えられる。

興味の学習の流れについては、基本的には既存手法と同じであるが、学習に用いるユーザの振る舞いに、視線が追加される。提案手法における興味の学習方法を表 1 に示す。

ユーザは注視した単語から対象記事に興味があるかどうかを判断していると考えられるため、興味の学習に用いる単語を、実際にユーザが注視していた単語に限定することで、より実際のユーザの実際の興味を学習できると考えられる。

3 評価実験

3.1 評価用システムの構成

視線を考慮した学習アルゴリズムの評価に用いるシステムの構成図を、図 1 に示す。

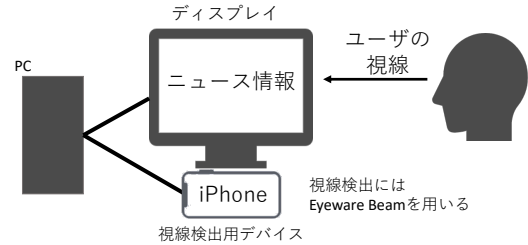


図 1: 提案手法評価用システムの構成図

本システムは、画面上にインターネットニュースを表示し、そのニュースに対するユーザの振る舞いと視線情報を取得する。ユーザに提示する画面は、記事見出し文を一覧で表示する画面と、任意の記事の詳細情報を表示する画面の 2 種類である。記事見出し文表示画面では、ニュースの見出し文が一覧表示され、各記事の右側に配置されるボタンをクリックすることで、該当記事の記事詳細画面が表示される。そして、記事見出し文表示画面では、複数の見出し文に対する取捨選択の際の視線移動を取得し、詳細情報の画面では、特定の内容を説明する長文に対する視線移動を取得する。このシステムにより、ユーザがニュースに目を通す時の視線の動き方を調べる。

3.2 実験方法

評価実験では、Yahoo!ニュースにて配信されている 9 カテゴリーの中から、あらかじめ 4 カテゴリーを選択してもらい、選択されたニュースカテゴリのニュースをユーザに提供する。あらかじめカテゴリを絞ることで、ユーザが興味を持っている内容のニュースが含まれやすくなる。Yahoo!ニュースではカテゴリ毎に 50 件のニュースが配信されているため、一度の実験で 200 件のニュースに対するユーザの振る舞いを取得できる。この実験を 6 回行い、1 回目から 5 回目の実験で得られたユーザの振る舞いを基に興味の学習を行い、6 回目の実験データを用いてフィルタリング性能の検証を行う。

フィルタリング性能の評価指標には情報 S/N 比、除去率、再現率の 3 項目を用いる。情報 S/N 比は、記事情報を選別する際にユーザが興味を持つと判断された記事情報の中でユーザが実際に興味を持った記事情報の割合を示す。除去率は、実際にユーザにとって興味のない記事情報の中で、記事情報の選別の際に「興味なし」と判定された記事情報の割合を示す。再現率は、実際にユーザにとって興味のある記事情報の中で、記事情報の選別の際に「興味あり」と判定された記事情報の割合を示す。

3.3 視認時間の調査

実験で得られた各ユーザの視認データから、文字情報に対する視認時間を調査する。各ユーザが詳細情報

Use of Gaze Information for the Purpose of Analogizing Interest in a Text

†T. Ogawa, K. Endo, S. Kobayashi

Graduate School of Science and Engineering, Ehime University

表 1: 提案手法における興味の学習

	詳細情報を表示する	詳細情報を表示しない
記事の最後まで目を通した	見出し文のすべての単語に対して興味あり学習を行う (従来手法と同じ)	見出し文のすべての単語に対して興味なし学習を行う (従来手法と同じ)
記事の途中まで目を通した	目を通した単語に対して興味あり学習を行う	目を通した単語に対して興味なし学習を行う
記事に目を通していない	操作ミスと判定し、学習を行わない。または、詳細文の目の通し方から興味の有無を類推する	学習を行わない

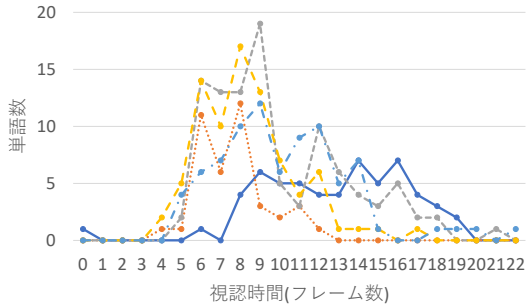


図 2: ユーザごとの単語視認時間のヒストグラム

を読む際の 1 単語当たりの視認時間のヒストグラムを図 2 に示す。この図では、横軸に視認フレーム数、縦軸に単語数を示す。本実験では、毎秒 30 回の頻度で視線情報を取得しており、視認時間はフレーム数で表す。

詳細情報を対象とした理由としては、詳細情報はユーザが知りたいと思表示した情報であるため、ユーザが関心を持って目を通していることが想定でき、流し読みではなくきちんと意味を理解しながら文字情報に目を通すときの視認時間が計測できると考えたためである。図 2 から、ユーザごとに 1 単語当たりの視認時間に差が出ていることがわかる。従って、ユーザごとに異なる基準で視認の有無を判別する必要があるといえる。

4 評価結果

提案手法の評価のために、既存手法との比較を行う。提案手法においては、図 2 の調査における各ユーザの同一視認時間の単語数が 2 語以上となった視認時間を、単語の意味を解釈した単語を判別する基準として用いる。また、ごく短時間であっても視認された単語を学習の対象とする場合についても評価を行う。

評価結果を表 2 に示す。表中の「視認手法」は、一度でも視認された単語を学習対象とする手法であり、「解釈手法」は、ユーザにより意味が解釈されたとされた単語を学習対象とする手法である。

表 2 から、再現率に関して割合が増加したユーザが多いものの、情報 S/N 比および除去率に関しては割合が減少したユーザの方が多いという結果になった。また、情報 S/N 比および、再現率に関しては、0%となっている箇所が多く、正確な評価が困難となっていた。その原因として、評価に用いたニュース記事に対して、それぞれのユーザが興味を持っているニュースの件数が少なかったことが挙げられる。表 3 にユーザ別の興味あり記事の件数とその割合を示す。なお、一度の実験で 200 件の記事を表示しているため、200 件の記事の

表 2: 提案手法の評価結果

ユーザ		情報S/N比			除去率			再現率		
		既存手法	視認手法	解釈手法	既存手法	視認手法	解釈手法	既存手法	視認手法	解釈手法
A		0.0%	0.0%	5.6%	98.5%	94.9%	91.3%	0.0%	0.0%	25.0%
B		0.0%	0.0%	0.0%	98.4%	97.9%	94.2%	0.0%	0.0%	0.0%
C		26.7%	20.7%	17.1%	94.0%	87.5%	84.2%	25.0%	37.5%	37.5%
D		9.1%	9.7%	7.1%	89.2%	84.9%	86.0%	14.3%	21.4%	14.3%
E		0.0%	7.7%	0.0%	93.7%	93.7%	91.5%	0.0%	9.1%	0.0%

表 3: 興味あり記事の件数と割合

ユーザ	興味あり記事数	割合
A	4	2%
B	9	5%
C	16	8%
D	14	7%
E	11	6%
平均	10.8	5%

うち興味のある記事の件数と割合を示している。先行研究では、少なくとも 1 割以上は興味のある記事が含まれていたことから、先行研究と比較すると、今回の実験では興味のある記事数が少なかったと言える。そのため、3つの評価指標を用いて評価を行ったとき、興味のある記事 1 つの影響が強いことが考えられる。

5 おわりに

提案手法の評価で用いた 3 つの評価指標では、提案手法と既存手法の比較を行うことは困難であった。そのため、今後は、興味あり記事件数の割合の影響の少ない評価の実施や、再実験などを行うことで、提案手法の評価を行う必要がある。

参考文献

- [1] 小森 健市, 西岡 豊, 柏木 紘一, 樋上 喜信, 小林 真也: “個人向け情報の配信を目的とした PUSH 型情報配信システムの構築”, 情報処理学会シンポジウム論文集, 第 2005 巻, 第 6 号, pp.249-252, 2005.
- [2] 西岡 豊, 柏木 紘一, 樋上 喜信, 小林 真也: “情報フィルタリングにおけるテキスト情報に含まれる品詞と受信者の特徴との関係-名詞と動詞に基づいた受信者の興味の類推-”, 情報処理学会シンポジウム論文集, 第 2004 巻, 第 7 号, pp.535-538, 2004.
- [3] 森 健, 柏木 紘一, 樋上 喜信, 小林 真也: “Ticker に対する表示操作履歴に基づいた興味の有無の推論”, 情報処理学会シンポジウム論文集, 第 2005 巻, 第 14 号, pp.111-116, 2005.
- [4] Takuya Ogawa, Keiichi Endo, Hisayasu Kuroda, and Shinya Kobayashi: “Analogy of the Degree of Interest in a News Article by Gaze Information”, Lecture Notes in Networks and Systems, Vol.255, pp.330-338, 2021.