

## ユーザの興味を喚起するブログ記事推薦システムの研究開発

北野光一<sup>†</sup> 寺口敏生<sup>†</sup> 田中成典<sup>‡</sup> 西江将男<sup>†</sup> 今城彰子<sup>†</sup> 佐藤香菜<sup>‡</sup>関西大学大学院総合情報学研究科<sup>†</sup> 関西大学総合情報学部<sup>‡</sup>

## 1. はじめに

ブログの増加[1]に伴い、ブログ記事の情報源としての価値[2]が高まっている。しかし、膨大な量のブログ記事の中から新しく関心を持った分野の情報を適切に検索することは困難[3]である。そこで、ユーザが過去に投稿したブログ記事から抽出した嗜好情報を用いて、ユーザの嗜好に合ったブログ記事を推薦する研究[4][5]が行われている。しかし、これらの既存研究では、ユーザの時系列的な興味の推移を考慮していないため、ユーザの最新の嗜好に合った情報を推薦できない問題がある。そこで、本研究では、ユーザが投稿したブログ記事の時間経過に伴う内容変化から興味の推移を抽出し、ユーザの最新の嗜好に合ったブログ記事の推薦手法を提案する。

## 2. 研究の概要

本研究では、ユーザの嗜好の変化を考慮したブログ記事の推薦手法を提案する。本システムの概要を図1に示す。本システムは、1) 学習機能と2) 推薦機能により構成される。入力データは、対象ユーザが記述したブログ記事群とし、出力データは、モデル・ユーザのブログ記事群とする。

## 2.1 学習機能

本機能では、カテゴリ別特徴ベクトルデータベースとユーザ別特徴ベクトルデータベースを構築する。まず、ブログ運営サイトからカテゴリ別とユーザ別にブログ記事を収集する。次に、カテゴリ別に収集したブログ記事に対して形態素解析を行い、カテゴリ毎に形態素の出現傾向を表した特徴ベクトルを算出し、カテゴリ別特徴ベクトルデータベースに登録する。また、ユーザ別に収集したブログ記事に対して形態素解析を行い、ユーザごとに形態素の出現頻度か

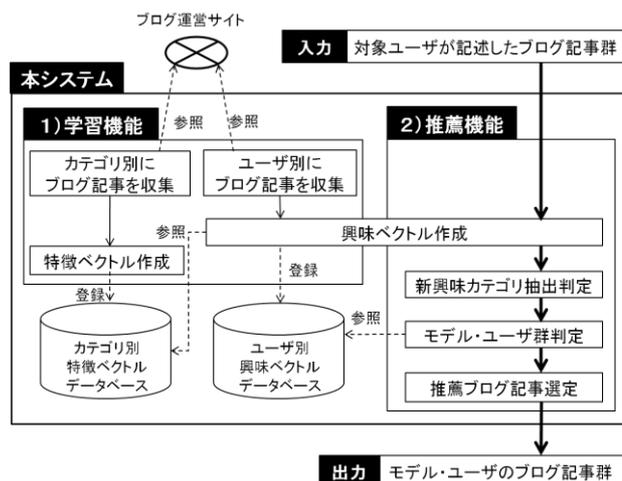


図1 本システムの概要

ら興味ベクトルを導出する。この時、ユーザの時系列的な興味の推移を考慮するため、時期毎にブログ記事中の形態素の出現傾向を算出し、カテゴリ別特徴ベクトルデータベースを参照して時期毎にユーザが興味のあるカテゴリを分析する。分析結果を基に、ユーザの興味ベクトルの中でも、長期的に興味を持ち続けているカテゴリに頻出する形態素の重みを上げる。そして、時間経過に伴って興味が減衰することを考慮し、過去に強い興味があったが、現在は興味が衰退しているカテゴリに頻出する形態素の重みを下げることで、各カテゴリにおけるユーザの興味度を考慮した興味ベクトルを作成する。最後に、その興味ベクトルをユーザ別興味ベクトルデータベースに登録する。

## 2.2 推薦機能

本機能では、まず、対象ユーザが記述したブログ記事群を入力し、学習機能における興味ベクトルの算出と同様の処理を経て、対象ユーザの興味ベクトルを導出する。次に、ブログ記事を投稿した時期を分析し、対象ユーザが最近興味を持ち始めた新たな興味カテゴリを抽出する。そして、ユーザ別興味ベクトルデータベースを参照し、対象ユーザの興味ベクトルとの類似度が高く、しかも、対象ユーザが新たに興味を持ち始めたカテゴリについて長期的に興味を示し

Development of System for Recommending Blog Articles Rousing Interest of User

<sup>†</sup> Koichi Kitano, Toshio Teraguchi, Masao Nishie, Akiko Imagi, Graduate School of Informatics, Kansai University, 2-1-1 Ryouzenji-cho, Takatsuki-shi, Osaka 569-1095, Japan

<sup>‡</sup> Shigenori Tanaka, Kana Sato, Faculty of Informatics, Kansai University, 2-1-1 Ryouzenji-cho, Takatsuki-shi, Osaka 569-1095, Japan

ているユーザをモデル・ユーザ群として判定する。最後に、モデル・ユーザ群のブログ記事で頻出している形態素を分析し、そのうち、対象ユーザのブログ記事に含まれていない形態素を含むブログ記事群を選定し出力する。

### 3. システムの実証実験と考察

本提案手法であるユーザの興味の変化を考慮したブログ記事の推薦手法の有用性を検証するために本提案手法である時系列を考慮する推薦システムと、単語の類似度のみを考慮し、時系列を考慮しない推薦システムの比較実験を行う。

#### 3.1 実証実験

本実験では、時間経過に伴う興味の変化を考慮することで、ユーザの最新の嗜好に合ったブログ記事を推薦できるかを検証する。まず、ユーザ1人が1つのブログを開設していると仮定し、Yahoo!ブログから60人分のブログ記事群を収集する。次に、収集したユーザのブログ記事群を用いて各ユーザが時期毎に興味のあるカテゴリを分析する。そして、ユーザがあるカテゴリに強い興味を示す前に投稿したブログ記事群をシステムの入力データとし、そのカテゴリに関連した内容を含むブログ記事を出力できるかを検証する。なお、ユーザが高い興味を示す分野を判別する指標として、対象となる興味分野に関する他ユーザのブログ記事をトラックバック機能により参照しているかどうかを用いる。評価指標には、正解データの割合を示す適合率を用いる。

#### 3.2 結果と考察

本実験では、ユーザがあるカテゴリについて興味を持ち始めた段階までに投稿したブログ記事群を用いて、その後ユーザが高い興味を示すブログ記事の内容を推測できるかを検証した。実験結果を表1に示す。表1より、本提案手法である時系列を考慮した推薦システムの適合率が0.67なのに対して、時系列を考慮しない推薦システムの適合率は0.43であった。本実験結果より、ユーザの最新の嗜好を取得するためには、単語の類似度のみではなく、時間経過に伴う興味の推移を考慮することが有用であることを証明した。しかし、実験結果を分析すると、流行に流されやすいユーザや興味が多岐に渡るユーザに対しては、適切な推薦対象を出力できないことが分かった。また、ユーザの興味が推移する過程は、ブログ記事に書かれている内容だけではなく、他ユーザのブログ記事をいつ閲覧したかといった履歴情報を考慮する必要がある。この問題に対しては、ユーザがブログ記事を閲覧した情報を分析することで克服できると考えられ

表1 実験結果

	時系列を考慮する 推薦システム	時系列を考慮しない 推薦システム
適合率	0.67 (40/60)	0.43 (26/60)

る。また、本研究では単語情報のみを解析対象としたため、ブログ記事中に存在する静止画像や動画画像といったユーザの興味を左右する情報を考慮できていない。そのため、「旅行」や「写真」といったユーザの視覚情報に依存する可能性の高いカテゴリにおいては、適切なブログ記事を推薦できない問題がある。この問題に対しては、カテゴリ別に静止画像や動画画像の出現率を分析することで、各カテゴリの視覚情報の重要度を定量化することに加えて、ブログ記事の本文だけでなく、HTMLタグにおけるalt要素を解析対象とすることで克服できると考えられる。

#### 4. おわりに

本研究では、時間経過に伴う興味の変化を考慮することで、ユーザの最新の嗜好に合ったブログ記事の推薦手法を提案し、実証実験により、その有用性を実証した。したがって、本提案手法は、ユーザが興味のあるブログ記事に到達するための一助となり、ブログの利便性向上に寄与できると考えられる。今後は、被験者実験を行い、ユーザの最新の嗜好を考慮することがユーザの興味の喚起において重要であることを検証する予定である。

#### 参考文献

- [1] Nardi, B., Schiano, D., Gumberecht, M. and Swartz, L. : Why We Blog, Communications of the ACM, ACM, Vol.47, No.12, pp.41-46, 2004.11.
- [2] Rubin, V. and Liddy, E. : Assessing Credibility of Weblogs, Proceedings of the AAAI Spring Symposium Approaches to Analyzing Weblogs (CAAW), the Association for the Advancement of Artificial Intelligence, pp.187-190, 2006. 3.
- [3] Ohmukai, I. and Takeda, H. : Metadata driven Personal Knowledge Publishing, Lecture Notes in Computer Science, Springer, Vol.3298, pp.591-604, 2004.11.
- [4] 森本和伸, 林貴宏, 尾内理紀夫 : MineBlog : 興味発見を支援する blog 記事推薦システム, 情報処理学会論文誌, 情報処理学会, Vol.47, No.4, pp.1171-1175, 2006.4.
- [5] 岸田真和, 倉本到, 渋谷雄, 辻野嘉宏 : ウェブログにおけるユーザの興味のあるエントリの推薦方法, ヒューマン情報処理研究会技術研究報告, 電子情報通信学会, Vol.104, No.746, pp.25-30, 2005.3.