

読書傾向分析に基づく書籍検索と推薦

藤本 政宏[†] 加藤 誠[†] 田中 克己[†]

[†] 京都大学情報学研究科 〒 606-8501 京都府京都市左京区吉田本町

E-mail: †{fujimoto,kato,tanaka}@dl.kuis.kyoto-u.ac.jp

あらまし 本研究では、ある特定の読書傾向を示す読者に基づいて書籍の検索を行う方法を提案する。既存の書籍検索システムでは「人気があること」以外の側面で書籍のランキングを取り扱うことが難しい。本研究では、Web 上に蓄積された読書記録の集合である読書傾向を集合知的な手がかりとして、特定の読者層によって検索結果の出力を変更する。提案手法では、検索結果の各書籍に対して特定の読者層に基づいたスコアを計算する。任意の本に対して、特定の読者層に基づく特徴を推定することによって、ユーザの読者層に基づいた検索を可能にする。

キーワード 情報検索, 情報推薦, データマイニング

1. はじめに

数多く出版されている本の中から、自身が読みたいと思う本を探し出すためには、検索システムや推薦システムの支援が不可欠である。ある特定の要求を持っている場合には検索キーワードを入力することで本を検索し、漠然とした要求を持つ場合には推薦システムによって推薦を受ける。しかしながら、どちらの場合においても、自分があまり読んだことのないような分野の本の中から、どの本を実際に読むか詳細に検討することは難しい。これは、その分野に関して多くの知識を持たないがために、本の紹介文や本のレビューなどを正確に理解できない可能性があるためである。また、Web などを通して本を購入する場合には、書店に比べて書籍の中身を閲覧することが難しく、オンラインでの書籍購入が一般化するにつれて、本の比較検討がより難しくなっている。

そこで本論文では、通常の実験クエリに加えて読者層を入力し、その読者層が主に読むような書籍を検索する方式を提案する。たとえば、「機械学習」という検索クエリに加えて、「プログラマ」という読者層を入力することで、プログラマが主に読むような機械学習の本（例えば、機械学習手法の実践に重点を置いたような本）を検索することができる。この方式の利点として、

- (1) ユーザ自身を包含するような読者層を入力することで自身の関心・知識に適した本を検索できること
- (2) 検索キーワードでは表現できないような、本の内容に関わる性質、例えば、専門性や実用性、新規性を暗に指定することができること
- (3) 書籍情報には含まれない、本の読者層という新たな側面から検索を行うことができること

が挙げられる。この方式によって、書籍に付随する紹介文やレビューといった理解に専門性を必要とする情報を用いずに、読者層を介して本の比較検討を支援することが本研究の目的である。

本論文では、特に、ある検索クエリの結果に含まれるような

本によって読者層を指定する方式について検討を行う。この方式においてユーザは、ある読者層が主に読んでいるような種類の本を検索結果として含むようなクエリを入力することで、読者層を指定する。例えば、「プログラマ」という読者層を指定するために、「プログラミング」に関する書籍が検索結果に含まれると期待されるような「プログラミング」や「Ruby」というクエリをユーザは入力する。提案手法は読者層を指定するクエリによって得られた検索結果から、それらをよく読むようなユーザを読者層として特定する。その後、通常の実験クエリによって得られていた検索結果を、指定された読者層が主に読んでいる順に並べ替えることで読者層による検索を実現する。

読者層による検索の有効性を検証するために、我々は読書 SNS から収集したデータを用いて実験を行った。本論文における実験では、あらかじめ用意したシナリオに応じてクエリおよび読者層を入力し、得られた結果が意図したとおりに並び替えられたかを nDCG@10 によるランキングの比較で検証した。

本論文の構成は以下の通りである。2 節では、推薦システムや書籍検索の関連研究について述べる。3 節では、提案手法について述べる。4 節では、データセットと予備実験について述べる。5 節では、まとめについて述べる。

2. 関連研究

本節では、本研究と関連する研究について言及し、本研究の位置づけについて述べる。

2.1 パーソナライゼーションと推薦システム

本研究は読書について集合知的な行動データを利用するが、こうした行動データをパーソナライゼーション (Personalization) に用いる研究は多く行われている。

パーソナライゼーションを Web 検索に適用するパーソナライズド検索 (Personalized Search) に関して、Jaime ら [1] は、検索タスクやクエリ選択、位置情報、性別や年齢、職業などの個人に関する情報について、その類似性に基いてユーザをグループに分類し、グループの情報を Web ランキングに利用する可能性について検討した。また Jaime ら [1] は、検索者の属するグループの情報を利用して Web ランキングを向上させる

パーソナライズド検索の手法を提案し、あるグループに基づくパーソナライゼーションが、そのグループに関連するクエリについて検索を改善させることを確認した。

乏しいクエリや曖昧な情報要求からでも身近な情報が得られるため、パーソナライゼーションはユーザの満足度を高める。その一方で、そのユーザの過去の行動やそのユーザの持つ社会的属性によって表示される情報が最適化されるということは、ユーザ過去の行動や状況に受け取る情報が影響を受けるということの意味している。E. Pariser はそうした状況をフィルターバブル (Filter Bubble) [2] と呼び、パーソナライゼーションには多様な視点や多様なコンテンツからユーザを遠ざける側面があることを指摘した。ACM RecSys2011 でこのフィルターバブルについて、推薦システムに対する影響に関してパネルディスカッション [3] が行われた。

本研究は、ユーザが自分の状況に制限を受けずに検索結果を受け取るように、パーソナライゼーションではない方法で読書における行動データを利用する。特定のユーザ集合 (グループ) に基づいてランキングを向上させるという点で、本研究は Jaime らの方法と似ているが、本研究がランキングに使用するグループは検索者の属するグループとは限らない。ユーザはランキングに用いるグループを自由に指定することができる。

2.2 書籍検索

SNS の情報を用いた複雑な書籍検索タスクに関して、CLEF [4] のソーシャル書籍検索 (SBS, Social Book Search) ラボの取り組みが挙げられる。ソーシャル書籍検索ラボは、複雑な書籍検索タスクにおいてユーザを支援する技術を研究・開発することを目的としている。ソーシャル書籍検索ラボは、書籍検索の複雑な性質に焦点を当てるため、伝統的で単純な検索シナリオとは異なるタスクを提案している。提案トラックでは、コンテキスト、ユーザのプロフィール、その他の関係性にもとづき、豊富な検索要求に対して書籍を提案するためのタスクを提供している。マイニングトラックでは、書籍からタイトルや著者をメタデータにエンティティ・リンキングするなど、ユーザをサポートするための自然言語処理に関するタスクを提供している。インタラクティブトラックでは、全く新しい情報検索のインターフェースを人に経験させ、その行動を調べることで、新しい検索インターフェースを提供するためのタスクを提供している。

本研究と同様に読書 SNS を用いて書籍検索・推薦を扱う研究もある。南ら [9] は適合フィードバックの枠組みによって、読書 SNS に存在するレビューに対するユーザのフィードバックから、そのユーザの観点を推定し、推定されたユーザの観点に基づいた書籍の推薦を行う方法を提案している。

南ら [9] の研究はレビューを考慮するが、我々の研究はレビューではなく読んだという事実そのものを考慮する。また、本を推薦させるユーザを選択するのに、投票ではなくクエリを入力を用いる点が前述の研究と異なる。ユーザの選択にクエリを入力を用いる研究としては、片岡ら [8] の Tweet 検索に関する研究が挙げられる。片岡らは適合フィードバックとして文脈クエリを入力して、ユーザを選択している。

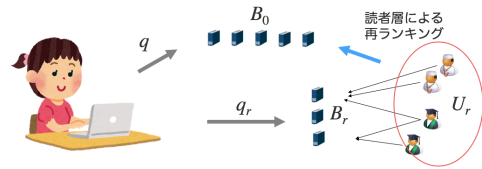


図1 提案システム概要

2.3 情報推薦

大量の情報が発信されるようになったことを背景として、どれに価値があるのかを特定するのを助ける道具として、情報推薦に関する研究が数多く行われてきた。

情報推薦の方法は内容ベースフィルタリングと協調フィルタリングに大別される。内容ベースフィルタリングは、検索対象のアイテムの内容を考慮して推薦を行う。一方協調フィルタリングでは、アイテムの内容を考慮することはせず、多くのユーザの行動履歴を集めた利用者データベースを用いて、行動履歴の類似性を用いて推薦を行う。

協調フィルタリングの代表的な方法としては Resnick ら GroupLens の方法が挙げられる [7]。GroupLens の方法では、アイテムを推薦する際、行動履歴がユーザと似ているデータベースのユーザ情報により大きな重みを与える。ここでは、すべてのユーザの好みは一貫しており、以前好きだったものは今後も好きである、という仮定が置かれている。

この仮定は前述のフィルターバブルの問題に関連する。そのためユーザが知らないアイテムを推薦することで、ユーザにとっての Novelty を向上させる協調フィルタリングに関する研究が清水ら [10] によって行われている。

3. 提案手法

本節ではまず、本研究の入力として扱う、読者層クエリについて説明する。その後、手法の概要を説明し、手法の各ステップの詳細について述べる。

我々は、知らない分野の書籍検索を支援するために、読者層によって検索結果の本を並び替えることを提案する。

しかし、読者層を直接提示することは、それを入力するユーザにとって負担が大きく現実的ではない。そこで、我々はユーザが読者層を直接入力するのではなく、「読者層を指定するクエリ」を通じて検索を行うことを提案する。このクエリを読者層クエリと呼ぶ。

ユーザは、対象とする読者層が読むような本に含まれるキーワードを指定することによって、読者層クエリによる読者層の指定を行う。たとえば「麻雀が強くなれるような理詰めた本」を探す際に、ユーザはまず「理屈っぽい人 (対象とする読者層)」に読まれる「麻雀」の本を探そうと考える。次にユーザは、対象とする読者層 (理屈っぽい人) が読むような本として「統計の本」を考え、そのような本に含まれるキーワードである「統計学」を、読者層クエリとする。

提案システム概要を図1に示す。提案システムでは、読者層クエリ q_r が与えられると、システムは読者層として読者の集合

を抽出し、その読者集合を用いて、検索結果をスコア付けする。このようにしてユーザは、読者層クエリによって、検索結果を並び替えることができる。すなわち、以下のステップからなる。

- (1) ユーザは通常クエリ q をシステムに入力する
- (2) 入力されたクエリから、システムの出力 B_q (初期ランキング) を得る
- (3) ユーザは読者層クエリ q_r をシステムに入力する
- (4) 入力されたクエリから、システムの出力 B_r を得る
- (5) システムは B_r から、読者層 U_r を取得する
- (6) $\forall b \in B_q$ それぞれについて、読者層 U_r に基づくスコア $\{r_{br}\}_{b \in B_q}$ を計算する
- (7) スコアにもとづいて B_q をランキングする
- (8) ユーザは再ランキング結果を確認し、 q_r を試行錯誤しながら、本を比較する

読者層クエリはひとつの通常クエリ q に対して、何度でも入力することができる。読者層クエリを入力するたびに、ステップ3以降の処理を行い、システムは B_q をランキングにしたものを出力する。

3.1 読者層の取得

読者層を、全読者集合 U の部分集合である U_r として定義する。ユーザ自身が読者集合を直接システムに入力することは負担が大きいため、本論文では、ユーザが読者層をクエリによって指定する場合を主に考える。

ユーザは書籍検索システムを利用する際に、通常クエリ q に加えて、読者層クエリ q_r を入力として与える。たとえば、「面白い数学の入門書が読みたい」というユーザの場合、「数学入門」が通常クエリ q 、「文春文庫」が読者層クエリ q_r となる。

U_r は次式によって表される。

$$U_r = \bigcup_{b_r \in B_r} U_{b_r} \quad (1)$$

ただし、 B_r は q_r による書籍検索システム (本論文では適当な書籍検索システムが利用可能であることを前提としている) の出力であり、 U_b は b の読者集合である。「文春文庫」が読者層クエリ q_r であるとき、読者層は「文春文庫を読んでいるユーザ」である。

3.2 読者層による本のスコアの計算

B_q のランキングを出力するために、 $\forall b \in B_q$ それぞれについて、読者層から計算されるスコア r_{br} を計算する方法が必要である。ここで、スコアは次の条件を満たすべきだと考える。

- (1) B_r の本をたくさん読んでいるユーザが読んでいるような本
- (2) U_r のなかでより多くの人々に読まれているような本

具体的な計算は次式によって行われる。

$$r_{br} = \frac{1}{|U_b|} \sum_{u \in U_r} \text{vote}(u, b, B_r) \quad (2)$$

$$\text{vote}(u, b, B_r) = \frac{1}{|B_u|} \sum_{b_0 \in B_r} \text{isRead}(u, b_0) \quad (3)$$

ここで、 $\text{vote}(u, b, B_r)$ はあるユーザが読んだ本の中で B_r が占める割合を計算する関数であり、本 b のスコアは $\text{vote}(u, b, B_r)$ が高いユーザがその本を読んでいればいるほど高くなる。

3.3 協調フィルタリングによる本のスコアの計算

各本のスコアの計算方法は上記で述べたとおりであるが、あるユーザが実際にどの本を読んだかという情報を用いるだけでは、さまざまな問題が生じうる。あまり多くの読者がいないような本に対しては、少数の読者から各本にスコアをつけるため、外れ値による影響が大きくなってしまふ。例えば、極端な場合、読者がいないような本はスコアを与えることができない。

そのため、我々は協調フィルタリングによって、あるユーザ u が読んだことのない本 b が将来的にどの程度読む可能性があるのかを推定し、読む可能性が高いような本もスコア付けの際に用いることを提案する。この方法によって、まだあまり読者がいないような本に対しても多数の潜在的読者によってスコアを算出することができる。

ユーザに対してアイテムを推薦する協調フィルタリングのアルゴリズムは、あるユーザとあるアイテムに対してスコアを与える。これを「ある読者がある本を読む度合い」と解釈し、読者層のスコアの平均をとることで、本を評価するのに用いることができる。

協調フィルタリングのアルゴリズムを用いたとき、ある読者層のある本への評価値は、次式で示される。

$$r_{br} = \frac{1}{|U|} \sum_{u \in U_r} \text{cf}(u, b) \quad (4)$$

ただし、 $\text{cf}(u, b)$ は、あるユーザ u がある本 b に対して与えるスコアである。

すべての読者、すべての書籍における読書記録をまとめた行列を $X \in \mathbb{R}^{|U| \times |B|}$ とする。行列の各要素は、要素に対応する u, b について、 u が b を読んでいる場合に 1、読んでいない場合に 0 をとる。以下、協調フィルタリングのアルゴリズムについて述べる。

3.3.1 メモリベースの協調フィルタリング

行列 X を直接用いる協調フィルタリングとして、Resnick ら GroupLens の方法があり [7]、 $\text{cf}(u, b)$ は次式のようになる。

$$\text{cf}(u, b) = \frac{\sum_{u' \in U} \text{sim}(u, u') |B_u \cap \{b\}|}{\sum_{u' \in U} \text{sim}(u, u')} \quad (5)$$

読者がある本を読んでいるかどうかは 2 値の状態しかとらないため、読者同士の類似度 $\text{sim}(u_i, u_j)$ は、読んでいる本集合の Jaccard 係数により求める。

$$\text{sim}(u_i, u_j) = \frac{|B_{u_i} \cap B_{u_j}|}{|B_{u_i} \cup B_{u_j}|} \quad (6)$$

3.3.2 行列分解

行列 X を直接用いない協調フィルタリングの方法として、行列分解 [5] を用いる方法がある。行列分解はこの非常に大きくかつ疎である行列 X を、より小さい行列に圧縮する方法である。行列分解では、行列の各成分は、 $|U|$ や $|B|$ よりもずっと小さい数の、 r 個の嗜好という潜在的な変数によって生まれて

いると考える。したがって、次式のように X を $Y \in \mathbb{R}^{|U| \times r}$ と $Z \in \mathbb{R}^{r \times |B|}$ を用いて、

$$X \simeq YZ \quad (7)$$

と分解する。 Y と読者の関係について、 Y からある行を取り出すと、ある読者に対応する r 次元のベクトルになる。ベクトルの各要素は、その要素に対応する読者の嗜好の強さと解釈できる。 Z と書籍の関係についても同様である。

行列分解の問題は、 E の各要素が 0 に近いことを表す誤差関数 L を最小化する Y, Z を発見する問題として定式化される。行列 E の各要素が 0 に近いほど、良い近似である。

$$E = X - YZ \quad (8)$$

このような Y, Z は解析的に得られない場合があり、その場合は何らかの誤差関数に基いて探索的に行列を求める。誤差関数としては、 E のフロベニウスノルム、KL ダイバージェンスなどが挙げられる。

3.3.3 非負行列分解

非負行列分解 (NMF, Non-negative Matrix Factorization) [6] は、非負行列に対して適用できる行列分解である。分解後の行列 Y, Z について、すべての要素が非負であるという制約を加えたくて、誤差関数を最小化する Y, Z を求める。

行列分解は、分解後の行列の要素からもとの行列の要素を復元する際に、線形和で復元を行う。非負行列分解では、分解後の行列の要素が負を含む場合とくらべて、分解後の行列の要素が非負であるとき、分解後の行列はより疎であることが実験的にわかっている。

3.3.4 行列分解を用いた協調フィルタリング

行列分解を用いた協調フィルタリングでは、読者と書籍それぞれの、低次元の特徴ベクトルのコサイン類似度をとる。したがって、 $cf(u, b)$ は次式のようになる。

$$cf(u, b) = \frac{\mathbf{u}^T \mathbf{b}}{\|\mathbf{u}\| \|\mathbf{b}\|} \quad (9)$$

ただし、 \mathbf{u} は Y から u に対応する行を取り出した縦ベクトルとしたもの、 \mathbf{b} は Z から b に対応する列を取り出した縦ベクトルとしたものである。

3.4 スコアに基づく検索結果のランキング

前小節で述べたスコア計算の方法により、 $\forall b \in B$ について実数値のスコアが計算できる。 B_q をスコアの昇順に並び替えることで、ランキングを出力することができる。

ランキングは、読者層クエリによって生成された読者層を反映したものになっている。たとえば、「面白い数学の入門書が読みたい」というユーザが通常クエリ「数学 AND 入門」、読者層クエリ「文春文庫」を入力した場合は、ランキングの上位であるほど「文春文庫を読んでいるユーザが読んだ、あるいは読みそうな数学の本」を表している。

読者層に基づくスコアでは、検索結果の各アイテムと読者層の。直接の読者の共通によってスコアの計算が行われる。ランキング上位には「文春文庫を”実際に”読んでいるユーザが読ん

でいる数学の本」がくる。レイティングの理由が明快であり、解釈しやすい一方で、”実際に”読まれていないかぎりはスコア付けが行われないので、ランキング中位～下位の微妙な差はわからない。

協調フィルタリングによるスコアの計算では、読者層に直接加わっていないユーザでも、ユーザどうしの類似度を通じてスコアに影響を与えるため、ランキング中位～下位の本であっても、非負なスコアを与えることができる。読者層に似ているユーザがその本を読んでいると、スコアが大きくなる。読者層に似ていないユーザがその本を読んでいるとスコアが小さくなる。これにより（読者層にも、読者層に含まれない人にも）広く読まれている本についてのスコアが小さく傾向にあると考えられ、売れている本のスコアを押し上げつける効果がある一方で、広く読まれている本のスコアが小さくなる可能性がある。

4. 実 験

本節では、提案手法を用いて行った実験とその評価について述べ、結果について考察を行う。本実験では、プロトタイプとして、提案手法に基づいた実際のシステムを構築し、その出力について考察する。

4.1 プロトタイプ

構築したシステムは、ユーザから通常クエリおよび読者層クエリを受け付ける。通常クエリによって、システムは検索結果を表示する。ここで、表示される結果は、Amazon.co.jp の検索結果とする。読者層クエリによって、(1)(2)(3) 式に基いてスコアを評価し、並び替えを行う。

4.2 データセット

読書系 SNS のひとつであるブックログ^(注1)から集めたデータを用いて実験を行う。読書 SNS は、ユーザが本を登録することができるソーシャル・ネットワーキング・サービスである。ブックログでは、ユーザは本を登録することで、Web 上に自分の本棚を作成することができる。ユーザと本の関係として、「読んだ」「読みたい」「読んでいる」「積読」「読みたい」という 5 つの状態を登録できる「読み終わった」に対しては、レビューを書くことができる。

本実験では、ユーザと本の関係として「読み終わった」「読んでいる」「積読」「登録」の 4 状態を分析対象とし、レビューは考慮しないこととする。収集されたデータは、本 492132 冊、読者 438503 人、登録数 14324418 件である。また、本 1 冊あたり読者数の中央値は 28.0 人、読者 1 人あたりの登録冊数の中央値は 6.0 冊である。なお、データセットはブックログの一部であることに留意する必要がある。

4.3 実験内容

ある検索タスクに対して各システムによって得られた検索結果を nDCG@10 で評価することにより、各システムの精度を評価した。本実験では、それぞれ 1 つの検索意図をもつ検索タスクを 25 個用意し、各検索タスクについて評価を行った。被験者は「～に役立つ本を探したい」「～であるような小説を探し

(注1) : <http://booklog.jp/>

表1 実験結果

ベースライン	提案手法	読者層クエリ	nDCG@10	
			ベースライン	提案手法
通常クエリ	通常クエリ			
SF	SF	脚本	0.1601	0.2746
IT 最新	IT 最新	採用担当	0.3409	0.3409
運動	運動	整形外科	0.2374	0.1444
健康食品	健康食品	薬剤学	0.4627	0.6717
がん	がん	基礎医学	0.389	0.697
インターネット 事件	インターネット 事件	判例	0.1682	0.0851
会社 制度	会社 制度	社労士	0	0
アイデア 特許	アイデア 特許	弁理士	0.474	0.3145
麻雀	麻雀	解析学	0.3219	0.435
問題 解き方	問題 解き方	解析学	0.3333	0.3333
プログラミング 入門 web	プログラミング 入門 web	ruby on rails	0.5479	0.2628
プログラミング 入門 ゲーム	プログラミング 入門 ゲーム	swift	0.3765	0.5506
料理 基本	料理 基本	一人暮らし	1.000	0.8744
曲 小説	曲 小説	ピアノ 教本	0.0212	0
数学 小説	数学 小説	解析学	0.832	0.8701
サイバー 小説	サイバー 小説	オライリー	0.8522	0.6521
戦争 小説	戦争 小説	軍事	0.8268	0.88
キリスト教	キリスト教	toeic	0.6728	0.6761
excel	excel	オライリー	0.2406	0.5978
アナログゲーム	アナログゲーム	ボードゲーム	0.5244	0.2884
電子回路 入門	電子回路 入門	Arduino	0.4067	0.7938
ボランティア 入門	ボランティア 入門	npo	0.4323	0.235
教える	教える	教育学	0.6537	0.3488
映画	映画	映画学	0.1297	0.1691
モノ デザイン	モノ デザイン	工芸	0.3566	0.4318
		Mean	0.4304	0.4371

たい」といった検索意図に基づき、各システムについてクエリを生成した。さらに、本実験の評価では、提案システムと既存システムで性能に大きな違いが現れた検索タスクについて、定性的な議論を行った。

4.4 実験結果と考察

実験結果を表1に示す。表1より、提案手法について定量的な精度改善は確認できなかった。本実験では、各手法の性能が検索意図によって大きく異なっている。

検索意図「Arduinoに役立つ電子回路の知識が学べる本」や検索意図「マニアックなExcelの使い方が載っている本」では、ベースラインに比べて提案手法の性能が大きく向上した。「電子回路入門」と検索した場合（ベースライン）と、さらに「Arduino」によって再ランキングを行った場合（提案手法）では、トップ6冊のうち4冊が変化していた。

- (1) 作る・できる/基礎入門 電子工作の素 (技術評論社, ISBN-10: 4774130788, 12位から1位)
- (2) 電子工作入門以前 (技術評論社, ISBN-10: 4774172847, 27位から3位)
- (3) Raspberry Piで学ぶ電子工作 超小型コンピュータで電子回路を制御する (講談社, ISBN-10: 4062578913, 13位から4位)
- (4) 回路シミュレータでストンとわかる!最新アナログ電子回

路のキホンのキホン (秀和システム, ISBN-10: 4798039411, 15位から5位)

(1) および(2)はレビューの評価が高く、検索の意図にも適合しており、役に立つことが期待できる。(3)はレビューの評価は高いものの、検索の意図には適合していない。(4)はレビューの評価が賛否両論だが、検索の意図には適合している。Amazon.co.jpから得られた各書籍の概要は次の通りである。

- (1) 電子工作にあたり、実際に使うための『道具』としての書物
- (2) 電気・電子の基礎理論や電子回路・部品の知識、マイコンの基本、プログラミングの基本
- (3) Raspberry Piで電子工作を学ぶ
- (4) いろんなアナログ電子回路の作りと動きがよくわかる

一方で、検索意図「Webアプリ開発のためのプログラミング入門本」では提案手法はベースラインに比べ性能が悪化していた。この検索意図においては、指定した読者層の傾向が、本のトピックにまで大きな影響を与えてしまい、入門者向きではない書籍が上位に来ていた。

このような検索意図において、精度の低下の主な原因としては、提案システムによる再ランキングが、既存システムが不適合だと計算した結果を上位に出力することが挙げられる。提案システムにおいても初期ランキングは既存システムによるもの

であり、初期ランキングはユーザが手に取るような本（すなわち、既存システムがユーザに対して適合を期待する本）を上位に、適合でない本を下位に置いている。そして、読者層クエリによって引き上げられた初期ランキングの下位にある書籍は、既存システムがユーザに対して適合を期待する本ではない。したがって、適合度の評価としては低下したものと考えられる。

5. ま と め

本稿では、読者層クエリによって書籍を並び替えるシステムを提案し、読者層による検索を可能にした。提案手法では、ユーザが理解しやすいような特徴にもとづいてスコア付けを行うことで、システムを利用するユーザが出力結果について理解しながら検索を行えるようにした。

また本稿では、読書 SNS ブクログから収集したユーザの読書傾向を用いて、提案手法の出力であるランキングについて評価実験を行った。評価実験の結果、提案手法による精度向上は確認できず、インタラクションを考慮した実験による再評価の必要性が確認された。

提案手法の課題としては、グラフデータのスパースネスへの対処が挙げられる。提案システムにおいては、あるアイテム集合を評価するユーザ集合は任意に選ばれる可能性がある。したがって、任意のアイテムについて任意のユーザが適切な評価値を与えられる方法の検討が必要であると考えられる。

提案システムの拡張としては、また、例えば複数の読者層クエリに基いてランキングを出力するといった、より複雑なクエリを入力するような拡張が考えられる。一方でこうした複雑な入力ユーザの利便性を損なうため、効果的なクエリの作成方法を検討する必要がある。

また現在は、単純なグラフデータを対象にしているが、他のソーシャルメディアの情報や、書籍のメタデータ、レビューテキストなどを含めてスコアを計算することが考えられる。

今後は、対象とする読書傾向データの拡充、実験デザインの再考、ベクトル空間法のような具体的なベースラインに基づく評価を行うことを予定している。

謝辞 本研究の一部は、文科省科研費基盤 (A)「多元的検索要求に対応できるオンラインデータマイニング検索方式の研究」(15H01718, 研究代表者: 田中克己), 文科省科研費若手 (A)「潜在的情報を引き出すための情報誘出技術に関する研究」(26700009, 研究代表者: 加藤誠) によるものです。ここに記して謝意を表します。

文 献

- [6] D.D. Lee and H.S. Seung, "Algorithms for Non-negative Matrix Factorization" Proc. NIPS 2000, pp. 556-562, 2000.
- [7] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, J. Riedl. GroupLens, "An open architecture for collaborative filtering of Netnews," In Proc. of the Conf. on Computer Supported Cooperative Work, pp. 175-186, 1994.
- [8] 片岡大祐, 加藤誠, 山本岳洋, 大島裕明, 田中克己. SNS グラフデータにおける文脈を考慮した適合フィードバック検索, 電子情報通信学会 第 9 回データ工学と情報マネジメントに関するフォーラム (DEIM2017) 論文集, 2017
- [9] 南大智, 牛尼剛聡. 書評 SNS におけるレビューの観点の違いを考慮したフィードバック型協調フィルタリング, 電子情報通信学会 第 8 回データ工学と情報マネジメントに関するフォーラム (DEIM2016) 論文集, 2016
- [10] 清水 拓也, 土方 嘉徳, 西田 正吾, 発見性を考慮した協調フィルタリングアルゴリズムに関する複数方式の検討", 人工知能学会論文誌, Vol.23, No.5, 330-343, (2008).

- [1] J. Teevan, M. R. Morris, and S. Bush. "Discovering and Using Groups to Improve Personalized Search. In Proceedings of the Second ACM International Conference on Web Search and Data Mining, " WSDM ' 09, pages 1524, USA, 2009. ACM.
- [2] E. Pariser. "The Filter Bubble: What The Internet Is Hiding From You." Viking, 2011.
- [3] Panel on The Filter Bubble. <https://acmrecsys.wordpress.com/2011/10/25/panel-on-the-filter-bubble/>
- [4] Social Book Search Lab, <http://social-book-search.humanities.uva.nl/>
- [5] Y.Koren, R.M. Bell and C.Volinsky, "Matrix Factorization Techniques for Recommender Systems," Computer, 42(8):30-37, 2009.