ユーザの嗜好を考慮したブックレビューランキング手法

若松　望† 常川　真央†† 松村　敦††† 宇陀　則彦†††

† 筑波大学情報学群知識情報・図書館学類　〒 305–8550 茨城県つくば市春日 1–2
†† 筑波大学大学院図書館情報メディア研究科　〒 305–8550 茨城県つくば市春日 1–2
††† 筑波大学図書館情報メディア系　〒 305–8550 茨城県つくば市春日 1–2
E-mail: †s0811659@u.tsukba.ac.jp, ††,†††{tsunekaw,matsumur,uda}@slis.tsukuba.ac.jp

あらまし　現在ブックレビュー投稿サイトに寄せられるレビューはあまりに大量であるため，レビューの有用性を判別し，ランキングする手法が研究されている．しかし先行研究の手法ではユーザの嗜好の個人差を考慮していない．そこで本研究では，ユーザ個人の嗜好を反映させるブックレビューランキング手法の実現を目的とし，ランキングアルゴリズムを提案する．個人の嗜好として，レビューの長さとレビューに含まれる単語の好みに着目した．比較対象として，ユーザ全員の平均的な好みを反映する手法を実装し，データセットを用いて比較評価実験を行った．その結果，提案手法は比較手法よりも上位での精度が高く，ユーザの参考になるレビューを上位に高い割合でランキングできた．
キーワード　ブックレビュー，ランキング，個人化

Book review ranking method considering user's preference

Nozomi WAKAMATSU†, Mao TSUNEKAWA††, Atsushi MATSUMURA†††, and Norihiko UDA†††

† College of Knowledge and Library Sciences, School of Informatics, University of Tsukuba
†† Graduate School of Library, Information and Media Studies, University of Tsukuba
††† Faculty of Library, Information and Media Science, University of Tsukuba
1–2 Kasuga, Tsukuba, Ibaraki, 305–8550 Japan
E-mail: †s0811659@u.tsukba.ac.jp, ††,†††{tsunekaw,matsumur,uda}@slis.tsukuba.ac.jp

Abstract　Currently, a great deal of book review is posted on the book review site. Therefore there are many studies on the methods of ranking reviews according to their helpfulness. However, these studies do not consider the difference in the users' preferences. In this study, we propose the book review ranking method considering user's preference. We focus on the length of book reviews and the terms included in the reviews as user's preference. To evaluate the proposed method, we compared the proposed method with the method using the average preferences of all users. As a result, the proposed method can rank helpful reviews higher than that ranked by the comparative method.
Key words　book review, ranking algorithm, personalization

1. はじめに

読書に関連した行動として，ブックレビューを読むということが広く行われている．これから買う本の情報を知るため，既に読んだ本について他の人の感想を知るためなどの様々な動機でブックレビューは読まれている．現在では Web 上に数多く存在するブックレビュー投稿サイトを通して，誰でも気軽に本の感想を投稿できるようになった．そのため，ブックレビュー投稿サイトのユーザは多くの人から寄せられたレビューを読み，参考にすることができる．しかし，投稿サイトに寄せられるレビューはあまりに大量であり，その中からユーザが参考になるレビューを見つけ出すことは困難である．そこで，多くのブックレビュー投稿サイトでは，本ごとにレビューのランキングを行う機能を導入している．レビュー投稿サイトで用いられているレビューランキング手法は，新着順と得票数（率）順に大別できる．

新着順では，投稿された日時の新しい順にレビューを表示する．投稿日時のみを基準としてランキングし，レビューが参考になるかどうかのフィルタリングは行わないため，ユーザが参考になるレビューをスムーズに探し出すことは困難である．

得票数順では，読んだレビューを参考にしたユーザの投票数の多い順にレビューをランキングする．得票率順は，同様にし

て，全票のうち参考になったという票の割合の高い順にレビューをランキングする．どちらにおいても多くのユーザの投票を必要とする．また，ユーザの投票を必要とするランキングでは，投票を行うユーザの心理にバイアスがかかることが指摘されている．Liu ら [1] は，Amazon(http://www.amazon.com/) のレビューランキングにおいて，(1) 他者の意見を見るときは肯定的に評価しやすい，(2) 多くの票を得ているレビューはさらに多くの票を得やすい，(3) 初期に投稿されたレビューが多くの票を集めやすいという 3 つのバイアスが存在することを明らかにした．これらのバイアスによって，ユーザにとって参考になるレビューであっても，現在用いられているランキングでは発見しづらい場合がある．

これらの問題を解決するために，レビュー文を分析してレビューをランキングする手法が研究されている [2] [3] [4] [5]．しかし，これらの研究ではレビューを読むユーザをひとつのまとまりとして扱い，個人の嗜好の差を考慮していない．本来ユーザは多様な嗜好を持ち，参考にするレビューも人によって違うということが考えられる．そのため，個人の嗜好を考慮することで，レビューランキングを改善できる可能性がある．

本研究の目的は，ユーザ個人の嗜好を反映させるブックレビューランキング手法の実現である．本稿では，個人の嗜好を反映させるランキングアルゴリズムの提案と検証を行う．

2. 関 連 研 究

Tsur ら [5] はレビューランキングアルゴリズム REVRANK を提案した．REVRANK は，「ある商品のレビューにはよく出現するが，一般の例文集にはあまり出現しない語は，その商品を最もよく表す重要な語である」という仮定をもとに，ある商品について書かれたレビューの集合から，重要な語の集合（Virtual Core レビュー）を生成する．次に，各レビューと Virtual Core レビューの類似度を計算し，類似度の高い順にレビューを並び替える．

Amazon のブックレビューデータを用いて REVRANK と Amazon のランキングを比較した実験の結果，REVRANK は，Amazon のランキングよりも高い精度で有用なレビューを上位にランキングできることが明らかとなった．しかし，本によっては Amazon とほぼ変わらない結果となることもあった．Tsur らは，好まれるレビューの長さはユーザや本によっても違うことを述べ，より精度を上げるためには長さの好みを考慮することが必要であるとしている．

異なる嗜好を持ったユーザが参考にする商品の評価について，今川ら [6] はレビュー投稿サイトを想定したシミュレーションを行った．想定するレビュー投稿サイトでは，ユーザは 2 つの評価をすることができる．1 つは商品に対する評価，もう 1 つは商品の評価を行ったレビュアに対するメタ評価である．2 つの評価値はどちらも数値で表される．レビュー投稿サイトのシステムは，商品の評価値とその商品の評価を行ったレビュアに対するメタ評価をもとに，商品のシステム評価値を求め，ユーザに提示する．システム評価値を提示する方法がユーザの満足度にどのように影響するかを提示方法間で比較した実験を行った．比較した提示方法は，(1) レビュアの平均的な評価値を提示，(2) メタ評価が最も高いレビュアの評価値を提示，(3) そのユーザにとってメタ評価の高いレビュアの評価値を提示の 3 つである．

その結果から，ユーザに商品の評価値を提示する方法として，そのユーザにとって評価の高いレビュアの評価値を提示する方が，全ユーザに同一の評価値を提示するよりも有効であることが示された．

そこで，本研究では，ユーザ個人の嗜好を反映させることでレビューランキングを改善する手法を検討する．個人の嗜好として，レビューの長さの好みとレビューに含まれる単語の好みに着目する．レビューの長さに着目した理由としては，Tsur らの手法の課題に長さの好みの反映があったことが挙げられる．また，レビューに含まれる単語の好みに着目した理由は，情報フィルタリングの分野で個人の嗜好として文書に含まれる単語が着目されてきた [7] ためである．Tsur らのランキング手法でも，レビューの有用性を決める要因として単語が用いられた．このことから，本研究でも，ユーザがそのレビューを参考にするか否かにレビュー中の単語が影響すると考え，ユーザの単語の好みに着目する．

3. 提 案 手 法

3.1 概　　要

本研究で提案するアルゴリズムは，レビュー投稿サイトにおいて，ユーザが過去に参考になるかどうか判定したレビューを利用できる状況を想定している．提案アルゴリズムは図 1 のように，好みの取得部分とランキング部分からなる．好みの取得部分ではユーザの判定済みレビュー集合から，レビューの長さの好み，レビューに含まれる単語の好み，これらの好みをランキングに反映させる最適な比重の 3 つを取得する．ランキング部分では，好みの取得部分で得られた長さと単語の好み，最適な比重を用い，新しく与えられたレビューのスコアを求める．

3.2 好みの取得部分

判定済みレビュー集合を用い，ユーザのレビューの長さの好みとレビューに含まれる単語の好みを取得する．以下，ユーザ u の判定済みレビュー集合を A_u，A_u のうちユーザ u が参考になると判定したレビューの集合を H_u とする．

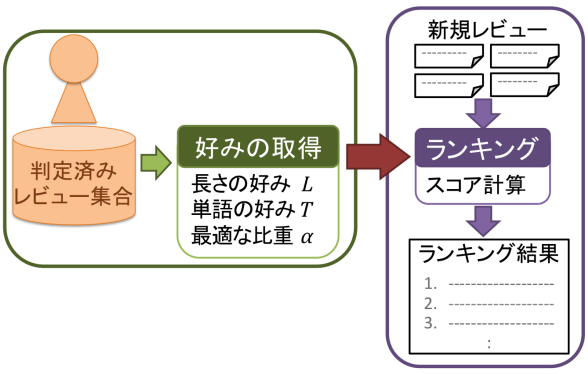

図 1　アルゴリズムの構成

3.2.1 レビューの長さの好み

ある長さのレビューをユーザがどれだけ好むかを測る．具体的には，レビューを長さによって階級に分け，階級ごとに (1) 判定済みレビューの件数，(2) ユーザが参考にした件数を求め，参考にされる割合を求める．階級 l の (1) の件数を $f_{A_u}(l)$，(2) の件数を $f_{H_u}(l)$ とすると，ユーザ u が階級 l のレビューを参考にする割合 $L(u,l)$ は以下の式で求められる．

$$L(u,l) = \frac{f_{H_u}(l)}{f_{A_u}(l)} \tag{1}$$

階級ごとに上記の割合を求め，ユーザの長さの好みを取得する．レビューの分布は文字数が多いほどまばらになることを考慮し，文字数が多くなるにつれ階級幅がログスケールで大きくなるようにする．各階級の端点を正の整数とし，文字数 c のレビューの階級を，

$$\lfloor \log(c) \rfloor \quad \sim \quad (\lfloor \log(c) \rfloor + 1) \tag{2}$$

とする．

3.2.2 レビューに含まれる単語の好み

ある単語をどれだけユーザが好むかを測る．具体的には，すべての単語について，(1) 判定済みレビュー全体での出現頻度，(2) ユーザが参考にしたレビュー全体での出現頻度を求め，参考にするレビューに含まれる割合を求める．単語 t の (1) の出現頻度を $tf_{A_u}(t)$，(2) の出現頻度を $tf_{H_u}(t)$ とすると，単語 t がユーザ u の参考にするレビューに含まれる割合 $T(u,t)$ は以下の式で求められる．

$$T(u,t) = \frac{tf_{H_u}(t)}{tf_{A_u}(t)} \tag{3}$$

判定済みレビュー中のすべての単語について上記の割合を求め，ユーザの単語の好みを取得する．

3.3 ランキング部分

好みの取得部分で求めたユーザの長さと単語の好み，これらの好みの最適な比重 α をもとにレビューのスコアを求める．ユーザ u にとってのレビュー r のスコア $Score(u,r)$ は，以下の式で求められる．

$$Score(u,r) = (1-\alpha) \times L(u,l_r) + \alpha \times \frac{\sum_{i=1}^{n} T(u,t_i)}{n} \tag{4}$$

$$(t_1,t_2,...,t_n) \in r$$

ここで，l_r はレビュー r の長さの階級，$t_i (i = 1,...,n)$ はレビュー r 中の異なる全単語を表す．新規に与えられるレビューそれぞれについて上記の式でスコアを求め，スコアの高い順にランキングする．

また，ユーザごとに，比重 α の最適化をする．判定済みレビュー集合を用いて，α の値を変化させてランキングを行い，ランキングの有効性を示す代表値が最大となる α の値を求める．ランキングの有効性の代表値は，正解数順位までの精度 [8] とした．正解数順位までの精度とは，ランキング中に全 n 件の正解データがある場合に，ランキングの上位 n 件までに含まれる正解データの割合を指す

表 1　選定した本のリスト

タイトル	著者
謎解きはディナーのあとで	東川篤哉
阪急電車	有川浩
下町ロケット	池井戸潤
夜明けの街で	東野圭吾
重力ピエロ	伊坂幸太郎
夜は短し歩けよ乙女	森見登美彦
スティーブ・ジョブス —人々を惹きつける 18 の法則	カーマイン・ガロ, 井口耕二 訳
もし高校野球の女子マネージャーが ドラッカーの『マネジメント』を読んだら	岩崎夏海
告白	湊かなえ
西の魔女が死んだ	梨本香歩

4. 評　　価

4.1 データセット

アルゴリズムの評価を行うために，個人による判定つきデータセットを作成した．データセットは，1 冊の本につき 100 件のレビューを 10 冊分（計 1000 件）用意し，14 人の判定者それぞれが参考になるかどうかを判定したものである．それぞれのデータは，(1) レビュアの名前，(2) レビュー本文，(3) 書かれた日付，(4) 判定者がそのレビューを参考になるかどうか判定した値からなる．

データセットに使用したレビューは，ブクログ (http://booklog.jp/) に投稿されたレビューから取得した．レビューを取得する本は，ブクログの本棚登録者数ランキング上位から 10 冊選定した．データセットに利用した本 10 冊を表 1 に示す．

各本につき 100 件のレビューはランダムに取得した．データ収集時は 2011 年 11 月 13 日である．取得したレビューデータから，レビュアとレビュー本文を本ごとに抽出し，判定用のレビューリストを作成した．さらに，14 名の判定者にレビューリストを見せ，それぞれのレビューがその人にとって参考になるかどうかを判定してもらった．判定は，(1) 参考になる，(2) 参考にならないの 2 値とした．

4.2 評価尺度

判定者をレビュー投稿サイトのユーザと想定し，データセット中のレビューのうち 9 冊分 (900 件) をそのユーザの判定済みレビュー，残り 1 冊分 (100 件) を新規に与えられたレビューとして，ランキングする．ランキングの評価尺度には以下の精度と再現率を用いる．判定済みレビューと新規レビューの組み合わせを変えて 10 回求め，その平均をとる．

$$\text{上位 } k \text{ 件での精度} = \frac{\text{上位 } k \text{ 件に含まれる正解データ数}}{k} \tag{5}$$

$$\text{上位 } k \text{ 件での再現率} = \frac{\text{上位 } k \text{ 件に含まれる正解データ数}}{\text{全正解データ数}} \tag{6}$$

4.3 比較手法

比較対象として，ユーザの平均的な好みを反映させたランキングを行うアルゴリズムを考える．比較対象のアルゴリズムで

は，個人の好みの代わりに，全ユーザの平均的な好みを用いる．

全ユーザの平均的な長さの好みと単語の好みを以下のように取得する．ユーザの集合を U とする．提案アルゴリズムにおける式 (1) と対応して，ユーザが階級 l の長さのレビューを参考にする割合 $L(U,l)$ を，以下の式により求める．

$$L(U,l) = \frac{1}{n}\sum_{i=1}^{n}\frac{f_{H_{u_i}}(l)}{f_{A_{u_i}}(l)} \quad (7)$$

$$(u_1,u_2,...,u_n) \in U$$

階級ごとに式 (7) の値を求め，ユーザの平均的な長さの好みとする．

提案アルゴリズムにおける式 (3) と対応して，ユーザが参考にするレビューに単語 t が出現する平均的な割合 $T(U,t)$ を，以下の式により求める．

$$T(U,t) = \frac{1}{n}\sum_{i=1}^{n}\frac{tf_{H_{u_i}}(t)}{tf_{A_{u_i}}(t)} \quad (8)$$

$$(u_1,u_2,...,u_n) \in U$$

単語ごとに式 (8) の値を求め，ユーザの平均的な単語の好みとする．

ランキング部分では，ユーザの平均的な好みをもとに，レビューのスコアを求める．ユーザ集合 U に対するレビュー r のスコア $Score(U,r)$ は，以下の式で求められる．

$$Score(U,r) = (1-\alpha)\times L(U,l) + \alpha \times \frac{\sum_{i=1}^{n} T(U,t_i)}{n} \quad (9)$$

$$(t_1,t_2,...,t_n) \in r$$

長さの好みと単語の好みをランキングに反映させる比重 α の値として，各ユーザの正解数順位での精度の平均値が最も高くなる α を求める．

5. 結　　果

5.1 精度・再現率

提案手法と比較手法でランキングを行い，それぞれ上位 k 件での精度・再現率を k の値を変えて求めた．精度の推移を図 2 に，再現率の推移を図 3 に示す．

精度を見ると，k の値が 5 のとき，有意水準 5%で提案手法と比較手法の間に有意差があった．ランキング上位での精度が高いことから，提案手法は比較手法よりも上位に高い割合でユーザの参考になるレビューをランキングできたとわかる．また再現率を見ると，k の値が 25 のとき，有意水準 5%で差があったが，それ以外の k の値 では提案手法と比較手法で有意差はなかった．

5.2 ユーザごとの正解数順位での精度

ユーザごとに，提案手法と比較手法それぞれの正解数順位 (k = 正解数) での精度を求めた．全ユーザの平均値とユーザごとの値を図 4 に示す．

全ユーザの平均値では，提案手法と比較手法の間で正解数順位での精度に有意差は見られなかった．しかし，ユーザごとに見ると，ユーザ B とユーザ J で提案手法と比較手法の精度に有意水準 5%で有意差があった．

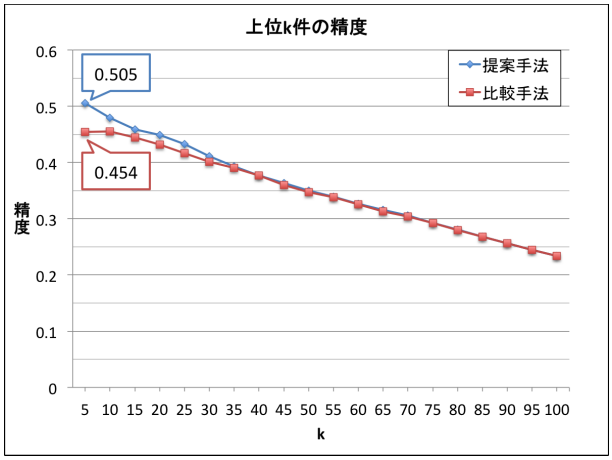

図 2　上位 k 件でのランキングの精度の推移

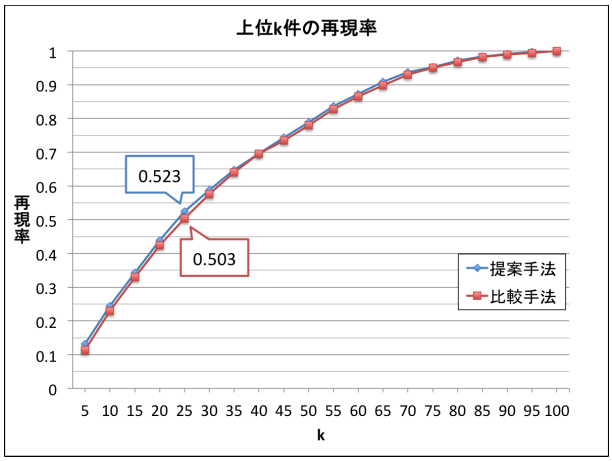

図 3　上位 k 件でのランキングの再現率の推移

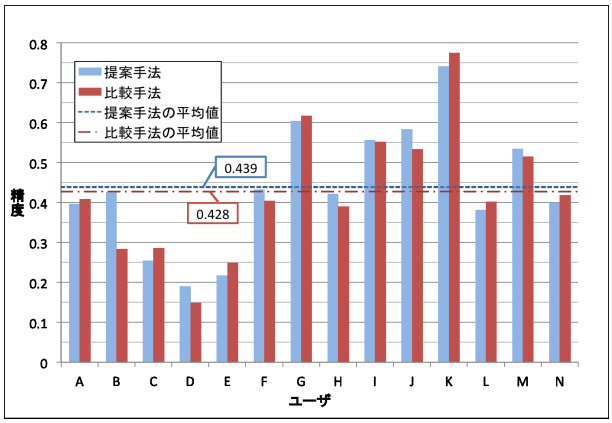

図 4　正解数順位での精度

6. 考　　察

6.1 ランキングの有効性

ランキングの有効性を測るために，正解数順位での精度の比較を行った．全ユーザの平均値を見ると，提案手法と比較手法の間に有意差はなく，提案手法が比較手法よりも有効であるとは言えなかった．これは，提案手法と比較手法のランキング精度の差が上位でしか見られず，比較を行う順位ではほぼ差がないためであると考えられる．

ユーザごとに見ると，ユーザ B，J のみ正解数順位での精度に提案手法と比較手法の間で差が見られた．ユーザ B，J はともに提案手法の方が精度が高い．これは，2 人の好みが平均的な好みとは異なっており，個人の嗜好を反映させる提案手法のほうが有効であったためである．平均的な好みとの違いとして，特に，長さの好みの違いが見られた．

提案手法または比較手法によって長さの好みを取得する際は，式 (2) によってレビューを文字数に応じた階級に分けて扱う．階級と文字数の対応を，表 2 に示す．

ユーザ B が最も参考にする割合が高いのは階級 8 のレビューであった．平均的な好みとしては，ユーザが階級 8 のレビューを参考にする割合は，階級 7，階級 6，階級 5 に次いで 4 番目となっている．ユーザ B の長さの好みは平均的な好みと異なっていたため，提案手法の方が有効であったと考えられる．最も参考にする割合の高いレビューが階級 8 のレビューであるユーザは，ユーザ B とユーザ F である．しかし，ユーザ B のランキング中の正解データ数が 100 件あたり 8.6 件であるのに対し，ユーザ F は 21.3 件であった．提案手法と比較手法の精度の差はランキング上位ほど大きい傾向にある．そのため，精度の比較を行う順位が上位であるユーザ B は提案手法と比較手法で有意差が見られるが，ユーザ F は差が見られなかったと考えられる．

平均的な長さの好みとしては，ユーザが最も参考にする割合が高いのは階級 7 のレビューであった．個人ごとの好みを見ても，階級 7 のレビューを参考にする割合が 1 番目または 2 番目に高いというユーザが多かった．その一方でユーザ J は，階級 7 のレビューを参考にする割合の高さが 3 番目となっている．ユーザ J の長さの好みは平均的な好みとは異なっていたため，提案手法の方が有効であったと考えられる．階級 7 のレビューを参考にする割合の高さが 3 番目であるユーザは，ユーザ J とユーザ D であった．しかし，ユーザ D の正解数順位での精度を見ると，提案手法と比較手法で有意差はなかった．これは各階級のレビューを参考にする割合のばらつきがユーザ中で最も小さく，階級ごとの好みの差があまり見られなかったためである．さらに，ユーザ D は正解数順位での精度が，提案手法・比較手法ともに全ユーザの中で最も低い．これは，個人の好みとして取得した長さと単語の好みが，ユーザ D のレビューの好みにあまり影響していないためであると考えられる．ランキング操作を 10 回行ったとき，ユーザごとに求められた最適な α の平均と標準偏差，また比較手法によって求められた α の平均と標準偏差を表 3 に示す．ここで標準偏差は，α の値が α の最適化を行う際に用いるレビュー集合によってどの程度ばらつくかを示す．

表 3 の標準偏差を見ると，ユーザ D は，長さと単語の好みをランキングに反映させる最適な比重が，最適化に用いるレビューによって大きく変わる．そのため，ユーザ D のレビューの好みは，長さと単語の好みまたはそのバランス以外の要因によって決まっている可能性がある．

一方，提案手法・比較手法ともに最も正解数順位での精度が高かったのは，ユーザ K であった．ユーザ K の正解数順位での精度が高かった理由として，ランキング中の正解データ数の影響が考えられる．ユーザ K の正解数は，レビュー 100 件あたり 53.7 件と 14 人中で最も多い．ランキングされるレビュー 100 件に対してユーザ K の正解データの数が多いために正解数順位での精度も高いということが考えられる．

表 2　階級と文字数の対応

階級	文字数
1	3 〜 7
2	8 〜 20
3	21 〜 54
4	55 〜 148
5	149 〜 402
6	403 〜 1096
7	1097 〜 2980
8	2981 〜 8103

表 3　α の平均値と標準偏差

ユーザ	平均値	標準偏差
A	0.135	0.022
B	0.245	0.026
C	0.23	0.141
D	0.521	**0.250**
E	0.145	0.047
F	0.35	0.158
G	0.27	0.14
H	0.115	0.022
I	0.235	0.114
J	0.41	0.05
K	0.043	0.026
L	0.131	0.048
M	0.18	0.122
N	0.205	0.015
比較手法	0.48	0.14

6.2 課　　題

提案手法は比較手法よりもランキング上位に高い割合で，ユーザにとって参考になるレビューを集めることができた．しかし，正解数順位での精度に有意な差があるとは言えず，提案手法は比較手法よりも有効性が高いとは言えなかった．よって，精度の向上は課題であると言える．

本研究では，レビューの好みとして長さの好みと単語の好みに着目した．しかしユーザによっては提案手法，比較手法ともに単語と長さの好みを反映させた効果が見られなかった．長さと単語以外にも，レビューを参考にするかどうかを決定づける要因がある可能性が考えられ，他の要因を検討する必要がある．

7. 結　　論

本研究では，個人の嗜好を反映させるブックレビューランキング手法の提案と評価を行った．個人の嗜好として，レビューの長さとレビューに含まれる単語の好みに着目した．比較手法として，ユーザ全員の平均的な好みを反映するランキング手法を実装し，評価用データセットを用いて精度の比較を行った．

その結果，提案手法は比較手法よりも精度が高く，ユーザの参考になるレビューを高い割合で上位に提示することができた．提案手法によってユーザごとに求められた長さの好み，単語の好み，最適な比重には違いが見られた．そのため，提案手法は比較手法よりも適切に個人の好みをランキングに反映できたと言える．ただし，人によっては提案手法，比較手法ともにランキングの効果が見られない場合もあった．

本研究により，個人の嗜好を反映することでブックレビューランキングを改善できる可能性が示された．また，レビューの長さとレビューに含まれる単語の好みが人によって異なることが明らかとなった．今後の課題は，ランキング精度の向上である．そのために，単語と長さ以外の好みの検討が必要となる．さらに，実際のブックレビュー投稿サイトで利用されているランキングに提案手法を適用した場合の評価を行うことも必要である．

文　献

[1] Liu, Jingjing; Cao, Yunbo; Lin, Chin Yew; Huang, Yalou; Zhou, Ming. “ Low-quality product review detection in opinion summarization ”. In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning(EMNLP- CoNLL), 2007, p.334–342.

[2] Turney, Peter, D.“ Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews ”. Proceedings of the 40th Annual Meeting of the ACL, 2002, p.417–424.

[3] Kim, Soo-Min; Pantel, Patrick; Chklovski, Tim; Pennacchiotti, Marco.“ Automatically Assessing Review Helpfulness ”. Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing(EMNLP), 2006, p.423–430.

[4] Ghose, Anynda; Ipeirotis, Panagiotis.“ Designing novel review ranking systems: predicting the usefulness and impact of reviews ”. Proceedings of the ninth international conference on Electronic commerce, 2007, p.303–310.

[5] Tsur, Oren; Rappoport, Ari.“ Revrank: A fully unsupervised algorithm for selecting the most helpful book reviews ”. In Proceedings of the Third International AAAI Conference on Weblogs and Social Media(ICWSM), 2009, p.36–44.

[6] 今川孝博, 川村秀憲, 大内東.“ マルチエージェントによるメタ評価を導入したレビュア創発モデルに関する研究 ”. 情報処理学会研究報告, 139(17), 2005, p.97–102.

[7] 福島俊一. Web サーチエンジンの基本技術と最新動向（下）最新技術. 情報管理, 46(7), 2003, p.436–445.

[8] Kando, Noriko; Kuriyama, Kazuko; Yoshioka, Masaharu.“ Overview of Japanese and English Information Retrieval Tasks (JEIR) at the Second NTCIR Workshop ”. Proceedings of the Second NTCIR Workshop on Research in Chinese & Japanese Text Retrieval and Text Summarization, 2001, p.73–96.

[9] 庭野正義, マッキン ケネス ジェームス, 永井保夫. 適合率と再現率を用いた Web ページランキングシステムの評価. 東京情報大学研究論文集, 14(1), 2010, p.1–10.