

重要トピックを考慮したレビュー文書セットの提示

平間 智史 横山 晶一

山形大学大学院理工学研究科情報科学専攻

{tth54615,yokoyama}@{st,yz}.yamagata-u.ac.jp

1 はじめに

ECサイトやレビュー投稿サイトの普及により、商品やサービスに対するレビュー文書を誰でも投稿・閲覧できるようになった。人気商品では数十～数百件以上のレビューが投稿されるため、全てのレビューを読むことは難しい。そこで、既存のいくつかの投稿サイトでは他の閲覧者の投票に基いたレビューの提示を行っている。例えば、Amazon.co.jp¹では「参考になった」「参考にならなかった」をユーザが投票することができ、「参考になった」割合をもとに提示順を決定している。これによって上位に提示された個々のレビューの質は高くなるものの、図1のように上位に提示されたレビュー中でトピックの重複や偏りが生じたり、多くのレビュアーが言及しているトピックでも上位のレビューに含まれない場合があるなど、必ずしも実際の評判が上位のレビューに反映されているとはいえない。

多くのユーザが上位に提示されたレビューしか閲覧しないことや閲覧負担の面から考えると、上位N件(数件から10件程度)のレビューで商品のおおまかな評判を把握できることが望ましい。

本研究では、提示するレビューの件数があらかじめ与えられたとき、ユーザ(閲覧者)満足度の高いレビュー文書のセットを提示することを目指す。その上で、ユーザによる投票情報を使用せず、レビュー文書集合中の潜在トピック比率から重要トピックを考慮することでユーザにとって有益なレビュー文書のセットを提示する手法を提案し、それがユーザ満足度の向上につながることを実験により検証する。

まとめると、本研究の目的は以下の3点である。

- N件のレビューを提示する条件の中で、出来るだけ有益なレビュー文書セットの提示
- 上記タスクにおいて、潜在トピック確率に基づく重要トピックを考慮した提示手法の提案
- 実験により提案手法がユーザ満足度の観点で有効であることを検証

¹<http://www.amazon.co.jp/>

そろそろ声優考えませんか

445人中351人が「参考になった」と投票…
個々の声優のイメージで作品に影響が出るのを嫌ってあえて俳優や芸人を使っていますが、ハウルしかり……

頼むから声優は本職の人にして欲しい

369人中288人が「参考になった」と投票…
トトロ等の本業の声優さんが、やっていた作品に関しては何倍も高かったのですが、最近の作品になって芸人の起用が増えてから……

US盤の購入を考えている方へ

34人中27人が「参考になった」と投票…
国内盤は高いからと安いUS盤の購入を考えている方へ。先日US盤が発売されましたが、ロスレス音声に違いがありますので御注意ください。国内盤はオリジナル日本語音声(ロスレス(DTS-HD MA))です……

図1: トピックの重複や偏りの例

2 関連研究

レビュー文書に関する研究は多く行われており [1]、特にアイテムに対する評価観点の抽出 [2] や観点ごとの評価を要約する研究 [3] が盛んである。これらは評価観点がある程度決まっている search アイテム [4] においては極めて有効である。一方、体験してみないと価値が分かりにくい experience アイテム [4] においては、評価観点が比較的曖昧であり、レビュー文書全体として読んで意味を持つ場合が多い(書評が代表例)。本研究は主に experience アイテムに関するレビューを対象とし、レビュー文書をそのまま提示する。

提示リスト内での重複や偏りに対する研究としては、推薦システムにおける推薦リスト内での多様化の研究 [5] が行われている。これは適合度ベースで生成された推薦リスト内で類似度の高いアイテム組がある場合に、片方をリスト外の低類似度のアイテムと入れ替えるものである。本研究はリスト内での多様化を目指す点で共通するが、リスト内での類似性ではなく、レビュー集合全体におけるトピックの重要性を考慮する点で異なる。

レビュー文書を扱う場合の問題として、虚偽スパムレビューによる攻撃がある [6]。これらは自社の商品

に肯定的なレビューを投稿したり、競合他社の商品に悪意を持って否定的なレビューを投稿するといったもので、虚偽のユーザー投票と組み合わせてスパムレビューを上位に提示させる攻撃が容易である。本研究の提案手法に対して攻撃を成立させるためには、レビュー文書集合全体のトピック分布が大きく変化する程の大量のレビューを投稿する必要があるため、スパム攻撃に対する頑健性が期待できる。

3 提案手法

基本的な方針として、レビュー文書集合中で多くのレビューで言及されているトピックを重要トピックと定義し、出来るだけ提示レビュー文書中に含まれるようにする。また、各重要トピックについて代表的な文書を選択することで提示レビュー文書間でのトピックの重複を少なくする。

レビュー文書の特徴として、以下が挙げられる。

- 1つのレビュー中で複数のトピックについて記述される場合が多い
- 固有名詞、表記揺れ等が多く出現する一方、特に固有名詞等は重要なキーワードになりやすい

これらを踏まえた上で、対応する手法として以下の手法を提案する。

3.1 潜在トピックに基づく手法

任意の1アイテムのレビュー集合に対し、トピックモデル LDA (Latent Dirichlet allocation) [7] を用いて潜在トピック情報を推定する。推定されたトピック情報に基づいて重要トピックを決定し、重要トピックを最も良く代表する文書を選択する。LDA を用いることで、文書中の複数のトピックの混合を表現でき、固有名詞等を自然な形で扱えることが期待される。LDA については 3.2 節で述べる。

処理の流れ

ある任意の1アイテムのレビュー文書集合から、提示レビュー文書セットを選択する処理の流れは以下のようになる。

1. 提示文書数 T 、潜在トピック数 K を与える
2. 各入力文書 $d \in \{1, \dots, M\}$ に対し形態素解析²を行い、内容語を抽出しストップワードを除外する

²MeCab ver.0.996,
<http://mecab.googlecode.com/svn/trunk/mecab/doc/>

3. 得られた単語頻度ベクトルをもとに、LDA によるトピック推定を行う。推定の結果として、各トピックに対する単語確率 $p(w|z_t)$ 、各文書に対するトピック確率 $p(z_t|d)$ を得る
4. 文書集合全体でのトピック確率の高いトピック順に T トピックを選択する
5. 選択した各トピック z_t について $p(z_t|d)$ が最大となる文書 d を選択する (他のトピックにおいて既に選択されている文書は除く)

3.2 Latent Dirichlet Allocation

LDA は、文書はある特徴を持った単語の分布であるトピックの混合分布から生成されると仮定する生成モデルである。ここで、 α, β はそれぞれ文書-トピック多項分布とトピック-単語多項分布に対するディリクレ事前分布のパラメータ (ハイパーパラメータ)、 N_d は文書 d の文書長 (単語数) としたとき、LDA の文書生成過程は以下ようになる。

1. ディリクレ分布 $Dir(\alpha)$ から、各文書 $d \in \{1, \dots, M\}$ に対して、単語分布 θ_d を生成
2. ディリクレ分布 $Dir(\beta)$ から、各トピック $t \in \{1, \dots, K\}$ に対して、トピック分布 ϕ_t を生成
3. 文書 d 内の各語 $w_{dn} (n \in \{1, \dots, N_d\})$ に対して、
 - (i) 多項分布 $Mult(\theta_d)$ からトピック z_{dn} を生成
 - (ii) 多項分布 $Mult(\phi_{z_{dn}})$ から単語 w_{dn} を生成

本研究では Gibbs サンプリングに基づくモデルの推定 [8] を行う³。

4 実験

本研究における提示レビュー文書セットの評価を、実際に被験者にレビューを提示するユーザ満足度評価のアンケートによって行う。

4.1 実験データ

ユーザ投票の情報が取得可能で、かつ experience アイテムのレビューとして、Amazon.co.jp からレビュー文書を収集した。収集条件と実験での使用データについて表 1 に示す。なお、50 字未満のレビューについては収集の時点で除外した。

³GibbsLDA++ ver.0.2,
<http://gibbslda.sourceforge.net/>

表 1: 被験者への提示実験での使用データ

カテゴリ	書籍
アイテム条件	50 字以上のレビュー 20 件以上の投稿
アイテム数	72
1 アイテムあたり 平均/最大/最小文書数	68.46/871/23
合計文書数	4929

4.2 実験で用いた手法

ユーザ投票に基づく手法 (投票法)

文書 d に対する「参考になった」投票数を $vote(d)_{GOOD}$ 、「参考にならなかった」投票数を $vote(d)_{BAD}$ としたとき、「参考になった」割合

$$V_d = vote(d)_{GOOD} / (vote(d)_{GOOD} + vote(d)_{BAD})$$

が大きい順に T 記事を提示する。

ハードクラスタリングに基づく手法 (クラスタ法)

提案手法における LDA によるトピック推定は、文書に対して多重にクラスタに属することを許すソフトクラスタリングに相当する。前述したように、1つのレビューに複数のトピックが含まれる場合のモデル化が期待される。比較として、文書が1つのクラスタに対応するハードクラスタリングによる提示を行う。

- 3.1 節で述べた提案手法と同様の前処理の後に、tf-idf による重み付けを行い入力とする
- クラスタ数 $k = T$ として Ward 法を適用し、クラスタを得る
- 各クラスタ $c \in \{1, \dots, k\}$ について、

$$d = \arg \min_{d \in c} \|d - \bar{c}\|$$

となる文書 d をクラスタの代表として選択する。ただし \bar{c} はクラスタ c の重心である。

潜在トピックに基づく手法 (提案手法)

3 章で示した提案手法である。本実験ではアイテムごとに総記事数が大きく異なるため、トピック数はアイテムごとに可変とした。また、ハイパーパラメータは $\alpha = 50/K$, $\beta = 0.1$ とし、イテレーション回数は 300 回とした。

表 2: 有益さの平均評価値

手法	有益さの平均評価値
投票	2.95
クラスタ	2.45
提案	3.06

表 3: 意外な情報を含む割合

手法	意外な情報を含む割合 (%)
投票	53.2
クラスタ	50.9
提案	52.8

4.3 実験条件

被験者は情報系の 20 代の学生、被験者数は 9 人である。提示文書数は $T = 4$ とし、各被験者に対してランダムに選択した 24 アイテムそれぞれについてレビュー文書セットを提示した。実際のレビューの閲覧状況に近づけるため、レビュー文書の読み方はおよその内容を把握する程度の短時間での流し読みとさせた。

レビュー文書セット以外に被験者に与えた情報は、アイテム名 (書名)、「商品の説明」の「内容紹介」のみである。なお、被験者に対して先入観を与えることを避けるため、同一の被験者に対する同一アイテム別手法での提示はしなかった。

被験者への設問として、ユーザ満足度の評価のために、提示レビュー文書セットごとにアイテムについての評判を知る上での評価値 (有益でない: 1 ~ 有益: 4) を回答させた。その他に、意外な情報が含まれるレビューがあれば選択させ、他にレビューの選択について感じた点を自由記述式で回答させた。

4.4 実験結果

各手法での実験結果について、有益さの評価について評価値の平均を表 2 に、意外な情報が 1 つ以上含まれたレビュー文書セットの割合を表 3 に示す。

有益さの評価値において、提案手法はクラスタ法と比較して高い評価値を、投票法に対しては同程度かやや上回る評価値を得た。一方、意外な情報が含まれる割合については手法間で有意な差は見られなかった。

「魂を揺さぶるスケーター」初のチャリティー…
先日の世界選手権の羽生選手のFPの演技に魂を揺さぶられた一人です。スポーツ観戦をしていて、これだけ心を動かされたのは久しぶりのこと。羽生選手の心のこもった演技に引き込まれ、思わず涙があふれてきました……

構成は青嶋ひろの氏です
……すべての写真がカラーで使われていて、そこは嬉しい驚きでした。本書のコンテンツや、羽生選手ご自身については他の方がレビューしてくださっている通りですが、拝読していて、覚えのある感覚にもしやと思い最終ページを練ったところ、構成担当は青嶋ひろの氏と判りました……

写真集+書籍。印刷がよくて写真が映える
……写真集+書籍という趣きだが、印刷がなかなかよく、……東日本大震災の影響など、いろいろな障害を乗り越えてきたことがわかる。震災後は10月のシーズン開始まで60回のアイスショーをこなしながら各地を転々として練習を続けたという……

羽生選手の頭の良さにびっくり
……あまり期待しないで読み進めたところ、彼の頭の良さにびっくりしました。客観的に、正確に、自分を見つめることができる。ベテラン選手並みに経験値が高い。失敗からすぐに改善点を見つけ、レベルアップにつなげる。あらゆる経験を成長の糧にしている……

図 2: 提案手法によるレビュー文書セット提示例

4.5 考察

図 2 に提案手法によって提示されたレビュー文書セットの中で高評価値となった例を示す。このアイテムに対しては43件のレビューが投稿されている。例中のレビューは投票法の場合は上から順に、5、10、28、35番目となるレビューであり、投票情報を用いた場合とは異なるレビューが提示できている。それぞれのレビュー文書を見ていくと、トピックの重複が比較的小さいことが確認できる。重要トピックとして選択されたトピックの単語確率が上位のものを見ると、生成確率が高い順に「演技」「構成・装丁」「震災の影響」「選手の人柄」に関するトピックと解釈できた。文書セット中にも重要トピックが含まれていることが確認できる。

一方、提案手法において評価の低い提示文書セットを分析したところ、トピック推定結果において比率上位が分かりにくいトピックで占められている場合が多かった。

評価の高い提示文書セットでは、トピックの解釈が比較的容易なものが多かった。結果として提示された文書それぞれの論旨が明確になり、高評価につながったと考えられる。解釈困難なトピックが推定された場合を分析すると、全体のトピックに対する上位トピックの比率が小さい場合が多かった。これらのことから、単純にトピックの推定が失敗している場合には提示がうまくいかない結論付けられる。

クラスタ法について、併合途中のクラスタを分析したところ、クラスタ法においては上位のクラスタ併合にかなり無理があるように見受けられた。また、クラスタからの提示文書選択についても、クラスタ重心に近い文書がトピックに対して代表性がある文書とはいえない場合が多く、論旨が不明確になったのも低評価となった理由だと考えられる。

また、提案手法においては個々のレビューの質については投票法より劣ると思われる場合が多かった。質が明らかに低いレビュー(罵倒や無関係の情報)が1つでも含まれるとユーザー満足度は低下するため、文書選択時にある程度のフィルタリング(投票法との組み合わせなど)を行うことも考えられる。

5 おわりに

本稿では、レビュー文書集合内での潜在的トピックの比率から重要トピックを考慮することで、有益なレビュー文書セットを提示する手法について述べた。また、実際に被験者にレビュー文書セットを提示する実験により提案手法の有効性が確認できた。

ただし、実験においては被験者数やアイテム数が少ないこと、提示文書数が $T = 4$ の場合のみであることから課題が残る。今後はより多くの被験者、アイテム数での検証を行いたい。

参考文献

- [1] Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, Vol. 2, No. 1-2, pp. 1-135, 2008.
- [2] Nozomi Kobayashi, Kentaro Inui, and Yuji Matsumoto. Extracting aspect-evaluation and aspect-of relations in opinion mining. In *EMNLP-CoNLL*, pp. 1065-1074, 2007.
- [3] Ivan Titov and Ryan McDonald. A joint model of text and aspect ratings for sentiment summarization. In *ACL08*, pp. 308-316, 2008.
- [4] Sylvain Senecal and Jacques Nantel. The influence of online product recommendations on consumers' online choices. *Journal of Retailing*, Vol. 80, No. 2, pp. 159-169, 2004.
- [5] Neil J Hurley. Towards diverse recommendation. In *DiveRS11*, p. 1, 2011.
- [6] Nitin Jindal and Bing Liu. Opinion spam and analysis. In *WSDM08*, pp. 219-230. ACM, 2008.
- [7] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, Vol. 3, pp. 993-1022, 2003.
- [8] Thomas L Griffiths and Mark Steyvers. Finding scientific topics. *PNAS, Proc. Natl. Acad. Sci. USA*, Vol. 101, No. Suppl 1, pp. 5228-5235, 2004.