

他人のコンテンツ評価を用いたユーザの嗜好推測に基づくコンテンツ推薦

鈴木健太 濱川礼

中京大学大学院 情報科学研究科

1. 概要

本論文はユーザの好むコンテンツの名前（例：本や漫画のタイトル）からそのコンテンツの他人のレビューを Web 上で取得し、そのレビューに類似している他のコンテンツをユーザに推薦する手法について述べる。

システムはあらかじめコンテンツのレビューを固定レビューと変動レビューに分類し、この 2 種類の情報を分析した結果をコンテンツの評価情報として蓄積する。システムはユーザの好むコンテンツの名前を入力として受け取り、その評価情報と、あらかじめ蓄積した他のコンテンツの評価情報との間でマッチングを行い、評価情報の似ているコンテンツをユーザに推薦する。この手法により、システムはユーザの好むコンテンツの名前を入力のみでコンテンツの推薦が可能となる。

2. 背景・目的

Amazon などで行われているコンテンツ推薦の協調フィルタリングでは、売れているコンテンツが優先的に推薦される傾向がある。しかし筆者の経験では好むコンテンツの売り上げを見た場合、そのコンテンツが必ずしも売れているコンテンツであるとは限らなかった。実際には売れていないコンテンツの中にもユーザの嗜好と合致するコンテンツが埋もれている可能性がある。そこで本研究はコンテンツのレビューに着目した。レビューに含まれているコンテンツの特徴がユーザの嗜好とマッチングすれば、協調フィルタリングから探しにくいコンテンツを取得できると考えた。

3. 提案手法

様々なレビューの評価情報を内容ごとに比較するために、文章をレビューしたユーザに左右されない固定レビュー（あらかじめコンテンツの固有キーワードを取得しておき、その固有キーワードを含む文章）とレビューをしたユーザによって変化する変動レビュー（感想などの評価者で変動する内容の文章）の 2 つに分類する。

各分類のレビューの文章の中で頻出する単語を抽出し、その単語を各分類の評価情報とする。

（図 1）システムは入力としてユーザの好むコンテンツの名前を受け取り、そのコンテンツの評価情報と一致する割合の高い評価情報を持つ、他のコンテンツをユーザに推薦する。

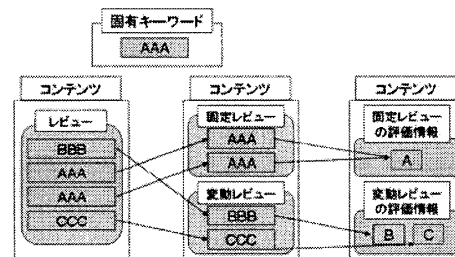


図 1 レビューから評価情報の取得

4. システムの全体構成

本システムは、あらかじめデータ収集を行うクローラと、そのデータを用いたマッチングの部分に分けられる。

● 固有キーワードの収集

コンテンツの固有キーワード（登場人物の名前や地名など）を Wikipedia よりあらかじめ収集しておく。

● レビューから評価情報の抽出

レビューの文章で、そのコンテンツの固有キーワードが含まれている文章を固定レビューとする。例えば A という登場人物がいる小説のレビューで「A の話し方が面白かった」という文章があった場合、A を含む文章なので固定レビューに分類する。そして残りの文章で、数字や記号のみの文章を除いた文章を変動レビューとして分類する。分類された文章の中から頻出する単語を抽出し、その単語をコンテンツのレビューの各分類の評価情報とする。

● ユーザの嗜好の設定

Amazon のカテゴリ分類によるジャンル検索か各ジャンル内でのキーワード検索で得た結果から、ユーザは好むコンテンツ名をシステムに入力する。そして蓄積されたそのコンテンツの評価情報をユーザの嗜好として設定する。

● コンテンツの推薦

ユーザの嗜好として設定した評価情報と、コンテンツの評価情報が一致する割合でコンテンツを推薦する優先度を決定する。

あらかじめ、システムに蓄積されたコンテン

ツの評価情報とユーザの嗜好の設定が一致する割合を固定レビューと変動レビューでそれぞれ取得する。(図2右)

ユーザはレビューの分類をシステムに入力し、あらかじめ取得した評価情報の一致する割合の情報から、システムはその分類でユーザの嗜好と評価情報が一致する割合の高いコンテンツをユーザに推薦する。(図2左下)

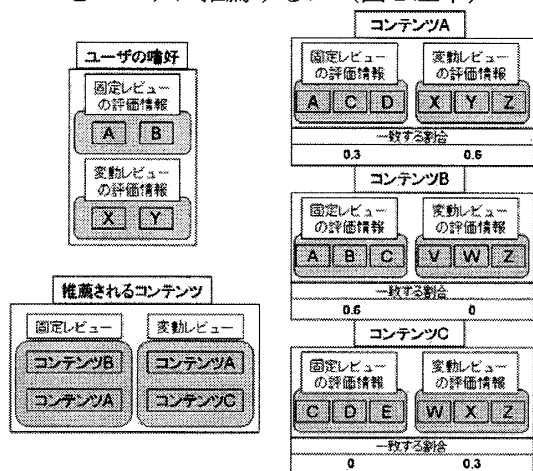


図2 コンテンツの推薦

5. 実験結果

システムにコンテンツ名を入力して評価情報を取得し、その評価情報から推薦されるコンテンツを取得する実験を行った。図3はAmazonの「世界の中心で、愛をさけぶ 小学館文庫(文庫)」のレビューから、変動レビューの評価情報を抽出した結果である。

この入力したコンテンツの変動レビューの評価情報と一致する割合の高い変動レビューの評価情報をもつコンテンツを、Amazonの「本 > 文学・評論 > 文芸作品 > 日本文学」のカテゴリ内の約25000冊の中からマッチングした結果を図5に、図4にAmazonでの推薦されるコンテンツの結果を示す。これらの結果からAmazonでは推薦されない特徴の似たコンテンツを取得できた。ここでAmazonからの結果の売り上げランキングの平均が78809位であったのに対し、本システムでは459475位と低い結果となった。

冬	時代	思春期	韓国	体験
白血病	象徴	性的	ドラマ	セックス
世間	理由	関心	病	母校
純愛	間違い	感情	婚約	最後
拍子	中学	ゴッチャ	恋愛	別れ
好き	冒険	相手	一昔	恋人
恋	高2	小説	結核	花道

図3. 変動レビューの評価情報

(世界の中心で、愛をさけぶ 小学館文庫(文庫))

タイトル
指先の花—映画『世界の中心で、愛をさけぶ』 律子の物語 (小学館文庫)
いま、会いにゆきます (小学館文庫)
そのときは彼によろしく (小学館文庫)
恋愛寫真—もうひとつの物語 (小学館文庫)
VOICE (アルファポリス文庫)
Separation—きみが選る場所 (アルファポリス文庫)
チャイ飲む?

図4. Amazonでのコンテンツ推薦

タイトル
少年の季節 (ノベル倶楽部)
性遍歴 (幻冬舎文庫)
あちらかばいッ (河出文庫)
天使と悪魔のあいだ〈下〉—さようならそしてこんにちは『片翼だけの天使』 (角川文庫)
国境の南、太陽の西 RMX (ダ・ヴィンチ・ブックス)
まわれ映写機 (幻冬舎文庫)
恋人よ〈下〉 (幻冬舎文庫)

図5. 変動レビューでのコンテンツ推薦

6. まとめと今後の課題

結論として、Amazonの関連検索からは取得できない売り上げランキングが低く、かつ特徴の似たコンテンツを得られた。この事により、本手法は特徴の似たコンテンツを発見する際には効果があることが判明した。

しかしコンテンツのレビューが無い場合はシステムから推薦されない問題点や、固有キーワードが取得できない場合そのコンテンツの評価情報が他のコンテンツと比べて偏ってしまう問題点がある。

よりユーザの嗜好に合ったコンテンツを推薦するために、今後はそれらのデータ補完やレビューの分類方法と評価情報のマッチングを改善する必要がある。

参考文献

[1] 土方嘉徳：嗜好抽出と情報推薦技術, IPSJ Magazine, Vol. 48, No. 9, pp. 957-965 (2007)
 [2] 河合由起子：嗜好情報に基づくニュースコンテンツの推薦とその応用, IPSJ Magazine, Vol. 48, No. 9, pp. 979-983 (2007)