

利用者の商品購入と Web ページ閲覧の履歴を用いた Web ページ推薦手法

山下真理子[†] 鈴木優[†] 川越恭二[†]

[†]立命館大学 情報理工学部

1 はじめに

実世界での人の商品購入履歴は、Web ページの閲覧履歴と共に購入者の嗜好を特徴的に表していると考えられる。なぜなら、人は商品を購入する際、自分の嫌いな商品や不必要な商品を購入する確率は低く、商品購入履歴には個人の嗜好が反映する商品が多くなると考えられるためである。しかし、商品購入履歴だけでは、商品と嗜好との関係が明確でないため、利用者の嗜好を抽出することが困難であるという問題点がある。そこで本研究では、Web ページの推薦精度を向上させるために、Web ページ閲覧履歴による Web ページ推薦システムに対して利用者の商品購入履歴を利用する手法を提案する。

具体例を用いて Web ページ推薦までの流れを説明する。利用者 A が“JAVA 入門”というタイトルで、“山田太郎”という著者の書籍を購入していたとする。まず、利用者 A の Web ページ閲覧履歴内を、JAVA 入門と山田太郎というキーワードを用いて検索する。検索結果の各 Web ページの特徴キーワードを抽出し、頻出キーワードが JAVA とアルゴリズムであったとする。この二つのキーワードを商品“JAVA 入門”という書籍を特徴付ける特徴キーワードとする。また、利用者 A が他にも JAVA という特徴キーワードを含む商品を多数購入していたとすると、特徴キーワード JAVA は利用者の嗜好を表すキーワードだと考えることができる。次に、利用者 B も JAVA という特徴キーワードを含む商品を多数購入していたとすると、利用者 A と B は嗜好が類似していると考えられる。このため、利用者 B の閲覧履歴内の Web ページを利用者 A に推薦する。

このように利用者が類似商品を購入していれば、同

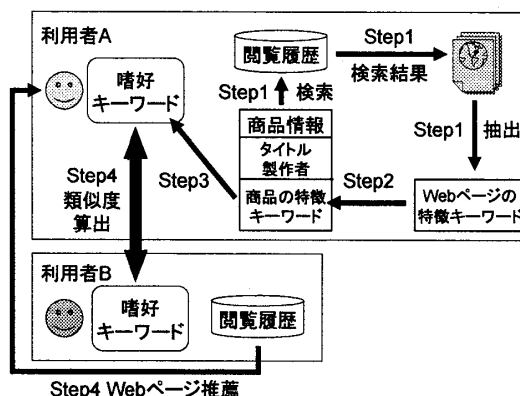


図 1: 提案手法の概要

一商品を購入していない利用者同士でも嗜好が類似していると判断できるため、推薦結果の Web ページがより利用者の嗜好に適合していると考えられる。

2 商品の特徴キーワード抽出手法を応用した Web ページ推薦

本研究では、利用者の Web ページ閲覧履歴から、利用者が購入した商品の特徴を表す特徴キーワードを抽出し、特徴キーワードを用いて利用者の嗜好を表す嗜好キーワードを導出する。そして、導出した利用者の嗜好キーワードから、利用者間の類似度を算出し、推薦対象の利用者と類似度の高い利用者の閲覧履歴内の Web ページを推薦する。

本研究では、利用者が購入した商品のタイトルや製作者をあらかじめ入手できると仮定する。

2.1 提案手法の概要

以下に、図 1 を用いて提案手法の概要を示す。

Step1 利用者の Web ページ閲覧履歴内を検索し、検索結果の Web ページから特徴キーワードを抽出する。詳細は 2.2 節で述べる。

Step2 Step1 で抽出したキーワードの中から、商品

Web page recommendation technique using the purchase history of goods and the access history of Web pages

Mariko YAMASHITA[†], Yu SUZUKI[†] and Kyoji KAWAGOE[†]

[†]College of Information Science and Engineering, Ritsumeikan University.

yamashita@coms.ics.ritsumei.ac.jp

の特徴キーワードを決定する。詳細は 2.3 節で述べる。

Step3 商品の特徴キーワードの中から、利用者の嗜好を表す嗜好キーワードを決定する。詳細は 2.4 節で述べる。

Step4 Web ページ推薦対象の利用者と、他の利用者との利用者間類似度を算出し、類似度の高い利用者の閲覧履歴を用いて Web ページを推薦する。詳細は 2.5 節で述べる。

2.2 Web ページの特徴キーワード抽出

利用者の Web ページ閲覧履歴内を、商品のタイトルや製作者をキーワードとして検索する。そして、検索結果の各 Web ページから、その Web ページを特徴付けるキーワードを抽出する。ここで、利用者の Web ページ閲覧履歴を用いる理由は、利用者の嗜好が特徴キーワードとして表れると考えられるためである。そこで、Web ページを特徴付けるキーワードを抽出するために、Web ページ内の文章を形態素解析する。形態素解析の結果から、名詞の単語だけを抽出し、TF-IDF 法を用いて単語の重みを決定する。各 Web ページごとに単語数が異なるため、重みの高い上位（例えば 2 割）の単語を、その Web ページの特徴キーワードとする。

2.3 商品の特徴キーワード決定

2.2 節で抽出した、商品に関する Web ページを特徴付けるキーワードは、商品の特徴付けるキーワードであるとも考えられる。このため、Web ページの特徴キーワードの中から、出現頻度の高い上位（例えば 10 個）のキーワードを商品の特徴キーワードとする。

2.4 利用者の嗜好キーワード決定

2.3 節で決定した、商品の特徴キーワードが同じである類似商品を利用者が多数購入していれば、その共通点は利用者の嗜好に関係していると考えられる。つまり、複数の商品に含まれる特徴キーワードは利用者の嗜好を表すキーワードであると考えられる。そこで、利用者が購入した商品の特徴キーワードの出現頻度を調べ、一定個数以上（例えば 5 個）の特徴キーワードを利用者の嗜好キーワードとする。

2.5 Web ページ推薦

2.4 節で導出した嗜好キーワードを、利用者間の類似度算出に用いる。まず、利用者間の嗜好キーワードが一致している度合いを利用者間類似度と定義する。この利用者間類似度を用いて、Web ページ推薦対象の利

用者と類似度が高い利用者を決定する。そして、推薦対象の利用者と類似度の高い利用者の Web ページ閲覧履歴から、推薦対象の利用者の嗜好に適合した Web ページを推薦する。

3 関連研究

高須賀ら [1] は、利用者の Web ページの閲覧履歴を用いて利用者間の類似度を算出し、推薦対象となる利用者と類似度が高い利用者の閲覧履歴内の Web ページを推薦する手法を提案している。本研究とは、利用者の行動履歴を用いて利用者間の類似度を算出している点で類似している。ただし、利用者の嗜好抽出で Web ページの閲覧履歴しか用いていない点で異なっている。

丹羽ら [2] は、利用者がソーシャルブックマークを登録する際に、付属することができるタグをクラスタリングし、利用者の嗜好導出にタグクラスタを用いる手法を提案している。そして、利用者に関係性の高いタグクラスタに含まれる Web ページを推薦する。本研究とは、タグと Web ページを関連付けて嗜好を抽出している点で類似している。ただし、利用者タグクラスタとの関連性を用いて Web ページを推薦する点が本研究と異なっている。

4 おわりに

本研究では、利用者の商品購入履歴と Web ページの閲覧履歴から、商品の特徴付けるキーワードを抽出し、利用者の嗜好導出に用いる手法について提案した。今後の検討事項を以下に述べる。まず、嗜好キーワードを用いた利用者間の類似度算出方法を検討する。次に、推薦対象の利用者の嗜好に適合した Web ページの算出方法について検討する。また、従来手法と本提案手法のシステムを作成し、Web ページ推薦結果を利用者による満足度によって比較する評価実験を行うことで、本提案手法の有効性を検討する。

参考文献

- [1] 高須賀清隆, 丸山一貫, 寺田実: “閲覧履歴を利用した協調フィルタリングによる Web ページ推薦とその評価”, 電子情報通信学会技術研究報告 [データ工学], Vol.107, No.131, pp. 115-120.
- [2] 丹羽智史, 土肥拓生, 本位田真一: “Folksonomy マイニングに基づく Web ページ推薦システム”, 情報処理学会論文誌, Vol.47, No.5, pp. 1382-1391.