

## 複数評価項目を考慮した推薦手法の提案

伊藤 ゆかり\*

波多野 賢治†

松本 尚宏‡

\* 同志社大学大学院文化情報学研究科

† 同志社大学文化情報学部

‡ 株式会社ワインスペース

### 1 はじめに

インターネットの普及により、通信販売を利用した商品の購入が一つの消費者行動として確立してきている。ユーザの嗜好に適した推薦を行うために、ユーザの嗜好の特徴を抽出しようとする研究では様々なアプローチがとられている [1]。しかしながら、その多くが「閲覧した」「閲覧していない」あるいは「評価した」「評価していない」といった二者択一の判断からの嗜好抽出が多い。

本稿では、複数の定量的な評価が付与された商品について、商品ごとにその評価項目の重要度を算出し、ユーザの付与した複数の評価項目を統合したデータからユーザの類似度を算出することで、商品へ付与されたデータから分かるユーザの評価の特徴を把握した推薦を行う手法を提案する。商品に付与される評価項目は、商品を表現する項目としてみると、商品の一部分に関する評価である場合や、全体としての評価である場合、評価項目同士に関連がある場合などが考えられる。また、どの評価項目を重要視して評価を付与するかも、ユーザによって異なる。このため、複数の評価項目を用いてユーザの類似度を測るには、それぞれの評価項目の重要度を考慮する必要がある。

評価として付与されたデータの数値差の総和をユーザ類似度とする場合では、表面上のデータに左右されて、本質的な類似度を測ることができない可能性がある。提案手法によって、アイテムに対する各評価項目の重要度を考慮することができ、より正確なユーザ類似度を測ることができる。

### 2 関連研究

ユーザの嗜好を抽出しようとする研究事例として、川前ら [2] は、アイテムに対する数値的な評価情報から類似の評価をしているユーザ同士を発見し、そのユーザ間で高評価のものを推薦対象ユーザに提示させている。ユーザの嗜好を区別するために、複数のユーザによって似た評価がされているアイテムは重要ではなく、

異なった評価がされているアイテムこそが重要であるとして、アイテムに重み付けを行い、対象ユーザとその他のユーザとの類似度を、アイテムごとの評価差の総和から求めている。

奥ら [3] による研究では、アイテムデータに付随する  $p$  次元特徴パラメータについて、ユーザの「満足」「不満足」の二択でパラメータの組合せについての嗜好を特徴付けている。これによりユーザの状況や気分が異なる場合でもその時々に応じた的確な推薦を試みている。

### 3 提案手法

複数の評価項目からユーザの類似度を算出するにあたり、本稿では、ユーザが重要視する評価項目の具体的な重要度を算出し、ユーザの評価がその重要度をふまえたスコアに置き換えられるようにするため、定量的なデータの特徴抽出手法として代表的な主成分分析を用いた。主成分分析とは、複数の変数間の相関関係を把握し、無相関の合成変数に縮約する手法である。各評価データに付与される主成分得点によって、全ての評価項目の重要度を考慮したユーザの評価傾向を表現した数値に置き換えることができる。

本システムの概要を、図 1 に示す。

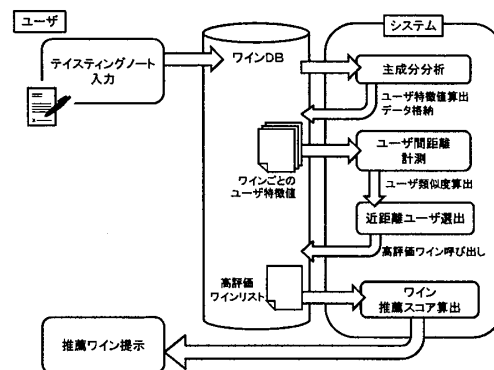


図 1: システム概要図

まず、ユーザから入力された評価データをデータベースに格納し、各ワインの評価データごとに主成分分析を行う。その結果、得られた各ユーザの主成分得点をデータベースに格納する。ワインに付与された評価項

Yukari ITOH\* Kenji HATANOf, and Takahiro MATSUMOTO‡  
 \* Graduate School of Culture and Information Science, Doshisha University  
 † Faculty of Culture and Information Science, Doshisha University  
 ‡ WineSpace Inc.

目はワイン雑誌として著名な「ワイン王国」\*で、ワインを紹介する際に用いられている評価項目を採用し、ボリューム、酸味、甘味、果実味、全体評価の、味に関する五項目を用いた。

次に、ワインごとの主成分得点から対象ユーザとその他のユーザとの類似度を算出して、その総和をユーザ類似度とする。ユーザ類似度は、ユーザ  $i$  を推薦対象者とした場合のユーザ  $j$  との距離  $L$  として算出するため、式 (1) のように定義した。  $n$  はワインの総数を表す。

$$L = \sum_{k=1}^n ((x_{ki} - x_{kj})^2 + (y_{ki} - y_{kj})^2) \quad (1)$$

ユーザ間の距離を計測できた後、距離スコアが 100 以下である上位のユーザを類似ユーザとして選出し、彼らの評価データから全体評価 4 以上を付与したワインを対象ユーザに提示する。

#### 4 評価実験

本稿における提案の実用性を検証するため、評価実験を行った。比較対象には、評価履歴が「有る」「無い」の合致数からユーザの類似度を算出して推薦を行う従来の推薦手法を用いた。提案手法を実装するために、計 14 名の被験者に 22 本のワインを試飲、評価させる試飲実験を行い、評価データを得た。

試飲実験から得られたアイテムへの嗜好評価を元に、推薦対象ユーザが高評価を付与したワインを正解として正解データセットを作成し、システムによる推薦結果の精度、再現率および  $F$  値を求めることにより評価実験を行った。実験結果を元に、提案手法と従来手法の精度、再現率、 $F$  値を表 1、11 点補完再現率精度曲線を図 2 に示す。

表 1: 11 点平均精度

	精度	再現率	F 値
従来手法	0.3147	0.638	0.4108
提案手法	0.3686	0.5775	0.4713

表 1 より、提案手法は、従来手法よりも精度、 $F$  値のスコアを上回る結果となった。しかし、図 2 より、0.4 を境として提案手法が従来手法を下回る結果となっている。これは、比較的上位で正解アイテムが推薦できた後、正解でないアイテムを多く推薦してしまっていることが問題となっている。この問題が起こる原因として、上位ユーザの選出を一律に距離値 100 以下とし

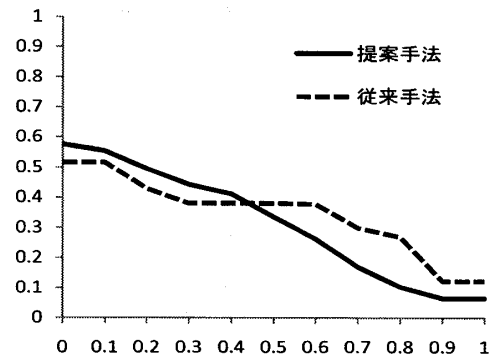


図 2: 再現率-精度曲線

ため、距離値のユーザ分布が考慮されず、類似でないユーザが選出されたと考えられる。距離値の設定はスコア分布を把握し、ワインごとに決定する必要がある。

#### 5 おわりに

本稿では、複数の定量的な評価からアイテムごとに評価項目の重要度を考慮し、全ての評価項目から総合的な評価数値を用いてユーザの評価傾向を特徴付け、推薦する手法を提案した。その結果、従来の推薦手法と比べてより高い精度でアイテムの推薦が行えた。

今後の課題として、考察における問題解決のために、ユーザの距離値分布状況を把握して距離値の境界を柔軟に決定することを考慮していく。また類似手法として、SVM による手法は、ユーザ分類を行う学習モデルとして最も認識性能が優れているとされている。主成分分析もユーザが増加するとスコアが変化することから、その変化の差異や有用性を検討するために、SVM との比較を行いたい。

#### 参考文献

- [1] 土方嘉徳. 情報推薦・情報フィルタリングのためのユーザプロファイリング技術. 人工知能学会誌, Vol. 19, No. 3, pp. 365-372, 2004.
- [2] 川前徳章. ユーザの潜在嗜好モデルに基づいた協調フィルタリング手法: ユーザの評価からユーザの嗜好の抽出. 電子情報通信学会技術研究報告, Vol. 103, No. 104, pp. 23-28, 2003.
- [3] 奥健太. ユーザコンテキストを考慮した情報推薦方式に関する研究. PhD thesis, 奈良先端科学技術大学院大学, 2009.

\*<http://www.winekingdom.co.jp/>