

多次元クラスタリングを用いたレコメンドシステムの研究

八 鍬 健太[†] 辻 秀一[‡]

東海大学大学院工学研究科[†] 東海大学電子情報学部[‡]

1. はじめに

近年、Web情報量は年々増加しており、従来のシステムでは対応できなくなっている。そこで現在のレコメンドシステムではシステムのより高度なパーソナライゼーションが進んでいる。そのパーソナライゼーションとは個人の好みや嗜好にあった情報を推奨するものである。本研究では k-means アルゴリズムを用いた多次元クラスタリングを使い、現在のシステムよりユーザに最も適合するよう事故学習機能を持ち、年齢や性別、住んでいる場所、検索履歴、好きなものなどの多数の事象に応じて調整された結果を出す事のできるレコメンドシステムの研究である。

2. 従来のレコメンドシステム技術

2-1 レコメンドシステムについて

レコメンドシステムとは、ショッピングサイトなどインターネットにおける情報提供サービスにおいて、不特定多数・特定個人の購買データやWebサイトへのアクセスデータなどから、人間では容易に発見できないパターンやルールを検出することにより、エンドユーザの嗜好や行動を推定し、ユーザのニーズに合致した情報（広告、商品、サービス情報等）を“推薦（レコメンド）”するシステムである。

2-2 クラスタリング技術について

クラスタリング (clustering) とは、分類対象の集合を、内的結合 (internal cohesion) と外的分離 (external isolation) が達成されるような部分集合に分割することである。統計解析や多変量解析の分野ではクラスター分析 (cluster analysis) とも呼ばれ、

いる。そして、そのクラスタリング技術には二基本的なデータ解析手法としてデータマイニングでも頻繁に利用されている2つの種類があり、階層的クラスタリングと非階層的クラスタリングといった種類がある。

今回はその中の非階層的クラスタリングの1手法である k-means 法 (k-平均法) を使用する。

その k-means 法とは、まずランダムに選んだ k 個のエレメントを初期クラスター中心 (seeds) とする。次に全てのエレメントを、クラスター中心との距離 (類似度) にもとづいていずれかのクラスターに属させて、各クラスターに属するエレメント間の距離 (類似性) より、そのクラスターのクラスター中心を再計算する。そして、クラスター中心の位置がほぼ定まり、収束判定条件を満たすまで直前の二ステップを繰り返すといったものである。

3. 提案方式

3-1 提案システム概要

従来のレコメンドシステムはユーザ情報と対象情報という情報で構成されたシステムであった。そのため、ユーザのダイナミックな状況変化に適応したレコメンデーションが出来ない問題があった。そこで今回の映画推奨システムでは個人情報、映画推奨情報の他に、映画鑑賞シチュエーションという第三的要素を加えた3次元での情報で構成し、ダイナミックな状況変化に適応したレコメンドシステムの提案を行う。

まず、本研究の提案システム全体図を Fig.1 に示す。本システムはクラスタリング技術のひとつである非階層的クラスタリングを使い、ユーザが望む商品を推薦するシステムである。全体システム概要は次のとおりである、まず情報提供者に個人情報として年齢、職業・好きなジャンルなどを登録する。そして、実際に推薦を行う場合は登録情報からの好きなジャンルのほかにユーザからいつ・どこで・誰とといった3つの項目を選んでもらい、その情報からユーザにとって見たい映画の推薦を行うシステムである。

Recommendation system with the method of multi-dimensional clustering

[†]Kenta Yakuwa

Graduate School of Engineering, Tokai University

[‡]Hidekazu Tsuji

School of Information Technology and Electronics, Tokai University

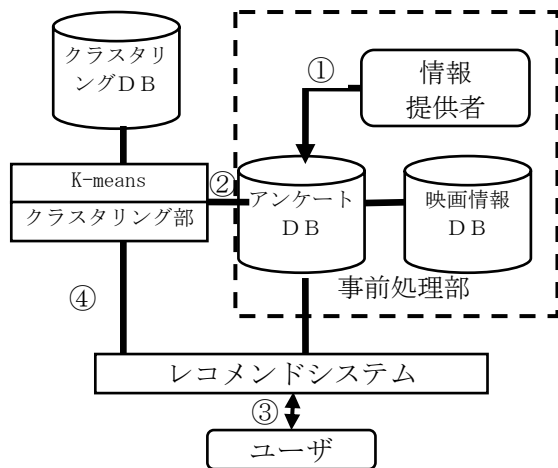


fig.1 システム全体構成図

3-2 詳細システムフロー

システムの詳細な流れは以下のようになる。

- ①アンケート情報を情報提供者から個人情報として、好みのジャンル、年齢、職業、第三的要素の映画鑑賞シーン（いつ、どこで、誰と）という情報を各映画ごとに収集を行って、情報をアンケートDBに収集する。この時、映画推奨情報を映画DBから各映画のジャンルや監督の情報は入手を行う。
- ②アンケートDB内のデータにクラスタリングを行なったあとに分類分けをし、クラスタリングDBへクラスタリングされたデータを保存する。この時行うクラスタリング処理は k-means 法を使ってシステムを構築する。手順としては以下のようなになる
 1. アンケートDB上にある映画情報を個人情報、映画推奨情報、映画鑑賞シーンのベクトル空間に配置を行う。
 2. ベクトル空間座標には以下の7要素を置き配置を行う。

<ul style="list-style-type: none"> ・ 個人情報 年齢・好きなジャンル・職業 ・ 映画推奨情報 ジャンル ・ 第三要素（映画鑑賞シーン） いつ・どこで・だれと

Fig.2 7つの要素区分

3. 今回は7要素を置くために7次元のベクトル空間から、非階層的クラスタリングの一手法である k-means アルゴリズムを使って、その空間で配置された要素のクラスタリングを行う。

③そして、ユーザが使う場合は個人情報を入力する。項目でいうと年齢、職業、好みのジャンル、そして推奨してもらいたい映画情報はどんなジャンルかである。次にその映画を見るとき第三要素（シーン）となる何時・何処・誰という3つをあわせた7つの項目で入力を行う。

④その情報から検出処理を行って、適切な情報（クラスター）をクラスタリングDBからユーザに出力を行う。クラスターの検出処理としてはユーザが入力した fig.2 にも書いてあるような7つの要素をベクトル空間に配置を行い、この座標をベクトル空間へ配置し、最も近い位置のクラスターを探し、ユーザへそれらのクラスターを出力する処理を行う。

4. 評価

従来のレコメンドシステムは個人情報と映画推奨情報といった2つの情報から推奨されていたものが、今回第三的要素というものを加えることにより、現在のレコメンドシステムの2次元情報から3次元情報へとより多くの情報量を使う事によって、より高度な推奨を行う事が出来た。今までのレコメンドシステムとは違い、ユーザの更なる好みを抽出する事が可能になったといえると考えられる。

5. 課題・考察

本研究では非階層的クラスタリング技術を使う事によって、多次元クラスタリングを用いたレコメンドシステムを行った。しかし、今後の改良としてファジィ c-means などのアルゴリズムでの実装などを行うことで、曖昧さ（複数のクラスターへ同時に所属する事）を加味する方法がある。また、今後のレコメンドシステムでは更なるパーソナライズ化が必要になってくるが、そのユーザが望む情報にいかにか近づける必要がさらに求められてくる。

6. 参考文献

<参考文献>

- [1] 高橋英史朗：「個人履歴情報を用いた Web 情報検索方式の提案」 情報処理学会第65回全国大会 / 3U-3、2003年3月。
- [2] Michael R. Anderberg, CLUSTER ANALYSIS FOR APPLICATIONS, ACADEMIC PRESS INC, 1973
- [3] マイケル J. A. ベリー, 「データマイニング手法」, SAS Institute Japan Ltd, 1999年