

関係性マイニングを用いた楽曲推薦手法の提案

| | |
|-----|---|
| 著者 | 比田 啓貴, 工藤 康生, 岡田 吉史, 村井 哲也 |
| 雑誌名 | 計測自動制御学会システム・情報部門学術講演会講演論文集 |
| 巻 | 2015 |
| ページ | 681-682 |
| 発行年 | 2015-11-18 |
| URL | http://hdl.handle.net/10258/3845 |

関係性マイニングを用いた楽曲推薦手法の提案

| | |
|-----|---|
| 著者 | 比田 啓貴, 工藤 康生, 岡田 吉史, 村井 哲也 |
| 雑誌名 | 計測自動制御学会システム・情報部門学術講演会講演論文集 |
| 巻 | 2015 |
| ページ | 681-682 |
| 発行年 | 2015-11-18 |
| URL | http://hdl.handle.net/10258/3845 |

関係性マイニングを用いた楽曲推薦手法の提案

○比田啓貴 工藤康生 岡田吉史（室蘭工業大学） 村井哲也（北海道大学）

概要 本研究ではラフ集合に基づく関係性マイニングを用いた楽曲推薦手法を提案する。工藤らが提案した関係性マイニングでは、2つの属性間での値の比較に基づく規則性も抽出することが可能である。本研究では、楽曲中でのコードパタンの出現回数比較に基づく特徴を抽出し、楽曲を推薦する手法を提案する。

キーワード: ラフ集合, 関係性マイニング, 楽曲推薦

1 はじめに

近年、Web上のオンラインショッピングサイトなどでユーザーの興味・嗜好にあった商品や情報を提供する推薦技術の研究が盛んに行われている¹⁾。そのためのデータマイニング手法としてラフ集合²⁾は有力視されている方法である。また、Kaneiwaら³⁾は系列データに対してラフ集合を使用する方法を提案している。本研究では、工藤ら⁴⁾による関係性マイニングが、2つの項目間での比較に基づく特徴を記述・抽出できることに注目し、関係性マイニングを用いて系列データのための推薦システムを構築することを目的とする。本研究では楽曲推薦を例に、提案システムの有用性を検討する。

2 ラフ集合

ラフ集合とは、大量のデータから決定表を作成し、対象の識別に必要な相対縮約を求める手法である。ラフ集合分析においてデータを示した表は決定表と呼ばれ、 $(U, C \cup D, V, \rho)$ の4対で定義される。Uは決定表の対象全体の集合、Cは条件属性の集合、Dは決定属性の集合、Vは各属性の属性値の集合、 $\rho: U \times C \cup D \rightarrow V$ はある対象のある属性に対して属性値を割り当てる関数である²⁾。また、決定属性により分割された対象の集合を決定クラスと呼び、対象がどの決定クラスに所属するかを識別するために最小限必要な条件属性の集合を相対縮約と呼ぶ。その属性値から対象が所属する決定クラスを識別するルールを決定ルールと呼ぶ。

3 関係性マイニング

ラフ集合による関係性マイニングは、属性間の相互関係を表す新たな属性を設定することで、2つの項目間の比較に基づく特徴をも抽出することを目的とする⁴⁾。

決定表 $(U, C \cup D, V, \rho)$ において、 $a, b \in C$ を任意の条件属性とし、Rを値の集合V上で定義された任意の二項関係とする。ある対象 $x \in U$ が存在し $(\rho(x, a), \rho(x, b)) \in R$ である時、かつその時に限り、属性aとbは関係Rの下で相互に関係するという。属性aの属性値 $\rho(x, a)$ と属性bの属性値 $\rho(x, b)$ との間に、二項関係が成立する対象 $x \in U$ の集合 $R(a, b)$ を、Rによるaとbの相互関係の支持集合と呼ぶ。R(a, b)は次式で定義される。

$$R(a, b) \stackrel{\text{def}}{=} \{x \in U \mid (\rho(x, a), \rho(x, b)) \in R\}. \quad (1)$$

定義した(1)式を用いて、相互関係決定表を作成する。相互関係決定表は $(U, C^R \cup D, V \cup \{0, 1\}, \rho^R)$ の4対で定義される。対象全体の集合Uと、決定属性の集合Dは元の決定表と同じである。また、集合 $R =$

$\{R_1, \dots, R_n\}$ をV上で定義された二項関係の集合とし、 C^R は相互関係集合を追加した新たな条件属性集合であり、次式で定義する。

$$C^R \stackrel{\text{def}}{=} C \cup \{aRb \mid R \in \mathcal{R}, R(a, b) \neq \emptyset, a, b \in C\}. \quad (2)$$

(2)式において、表現aRbは相互関係条件属性と呼び、aとbとの相互関係を記述した新たな条件属性である。V $\cup\{0, 1\}$ は元の決定表の各属性の属性値の集合に、相互関係条件属性の属性値の集合を加えたものである。 $\rho^R: U \times C^R \cup D \rightarrow V \cup \{0, 1\}$ は元の決定表と同様に対象と属性からなる属性値を割り当てる関数であり、 $\rho^R(x, aRb)$ を次式で定義する。

$$\rho^R(x, aRb) \stackrel{\text{def}}{=} \begin{cases} 1, & \text{if } x \in R(a, b), \\ 0, & \text{それ以外.} \end{cases} \quad (3)$$

ただし、通常の条件属性 $a \in C$ での属性値 $\rho^R(x, a)$ は $\rho^R(x, a) \stackrel{\text{def}}{=} \rho(x, a)$ と定義する。

相互関係決定表は、元の決定表に出現する条件属性間の相互関係を、相互関係条件属性を用いて付け加えた決定表である。

4 提案手法

本研究では、コードの系列として表された楽曲データから決定表を作成し、ユーザが指定した1曲をクエリとして用いて、ユーザが好むと思われる曲を推薦する手法を提案する。

4.1 楽曲データからの決定表作成

長久保の方法⁵⁾を改良し、楽曲データから決定表を作成する。本研究で扱う楽曲データ自体は属性や属性値を持たず、決定表を作成することができない。そのため、条件属性や決定属性を設定する必要がある。まず、条件属性を以下に設定する。

- 条件属性：楽曲データに表れる可変長の部分系列、
- 条件属性の値：対象の楽曲データで現れる部分系列の出現頻度。

本手法では部分系列を抽出する方法として、可変長N-gram法⁶⁾を用いる。条件属性とする部分系列の個数は対象数の増加に伴い大幅に増えていくため、各部分系列について、楽曲データ全体での出現頻度の合計に閾値を設定し、閾値以上の部分系列を条件属性とする。次に決定属性dを以下の手順で設定する。

1. クエリとしてユーザが好みの対象を1つ選択する。
2. 条件属性となる各部分系列について、クエリにおける出現頻度を調べ、出現頻度が多い順に3種類の属性 $a, b, c \in C$ を選択する。

3. 選択した属性 $a, b, c \in C$ をすべて含むかを表す属性 d を新たに作成し, 対象 $x \in U$ の属性 d での値 $\rho(x, d)$ を以下に定義する.

$$\rho(x, d) \stackrel{\text{def}}{=} \begin{cases} 1, & \rho(x, a) \neq 0 \text{ かつ } \rho(x, b) \neq 0 \\ & \text{かつ } \rho(x, c) \neq 0, \\ 0, & \text{それ以外.} \end{cases} \quad (4)$$

4. 作成した属性 d を決定属性とする.

以上より, 楽曲データから決定表を作成することにより, ラフ集合で用いることが可能となる.

4.2 楽曲推薦手法

ユーザの好む曲に基づく楽曲推薦手法を提案する. 本手法は以下の手順で行われる.

1. ユーザが好む楽曲を 1 曲選択してクエリとし, 4.1 節で述べた手法で楽曲データから決定表を作成し, 更に相互関係決定表を作成する.
2. 確率的選択を用いた縮約計算およびその並列分散化⁷⁾を用いて, できるだけ多数の縮約を相互関係決定表から抽出する.
3. 得られた縮約の中で評価値⁸⁾が最も良い縮約を選択し, 相互関係決定表から同値類を算出する. 同評価値の縮約が複数ある場合は, クエリを含む同値類に含まれる楽曲が最も少ない縮約を用いる.
4. 求めた同値類で, クエリを含んだ同値類からクエリ以外の楽曲をユーザに推薦する.

5 実験

提案手法の有効性を検証する実験を行った. 実験では, 長久保らがアンケートを用いて作成した, 270 曲の楽曲からなるデータセット⁵⁾を用い, 予め被験者 1 名が各楽曲に対して, 曲の評価を 5 段階評価 (1: 好まない~5: 好む) で与えた. 被験者が高評価 (評価値 5) とした 26 曲について, それぞれクエリとする楽曲として用いて, 楽曲データから相互関係決定表を作成した. 相互関係条件属性の作成に用いる二項関係 R は $>$ および $<$, $=$ とした.

また, 実験に用いるパラメータとして, 条件属性として用いる部分系列を選択するための, 出現頻度の閾値を 25, 並列分散化の試行回数を 10000 回とした. なお, 本実験では決定属性設定の際に選択した条件属性は, 相互関係決定表から取り除き, 選択した条件属性を含んだ関係性条件属性はそのままとして実験を行った.

6 結果と考察

実験結果を Table 1 に示す. Table 1 における, 推薦曲数および楽曲評価値, 縮約評価値の 3 項目は, いずれも推薦曲数が 10 曲未満の場合と 10 曲以上の場合における, それぞれの平均値を表す.

推薦結果をクエリごとの推薦楽曲数が 10 曲未満の場合と 10 曲以上の場合で比較すると, 10 曲未満の場合は, 10 曲以上の場合より平均評価値が若干高い値となった. これは推薦楽曲数が少ない場合ではユーザが好みの楽曲を推薦できる可能性が高いと考えられる. しかし, 平均縮約評価値の項目で, 10 曲以上推薦する場合の方が値が高いのは, 縮約の評価値が高いほど縮約属

Table 1: Experiment results.

| | クエリ | 推薦曲数 | 楽曲評価値 | 縮約評価値 |
|--------|-----|-------|-------|-------|
| 10 曲未満 | 14 | 4.42 | 3.40 | 0.14 |
| 10 曲以上 | 4 | 35.50 | 3.22 | 0.39 |
| 推薦不能 | 8 | — | — | — |

性の種類が減少するため, 同値類のパターンが少なくなり推薦楽曲数が増加しやすくなるためと考えられる.

また, 推薦楽曲数が 10 曲未満, 10 曲以上のどちらの場合も, 楽曲データ全体での楽曲の平均評価値 (3.05) と比べると評価値が高い. このことから本手法による推薦結果は, ユーザの好みがある程度反映できていると考えられる.

一方, クエリ 26 曲中の 8 曲で, 楽曲の推薦ができなかった. 本実験で推薦不能となった結果の内訳は以下の 2 通り存在した: 1) クエリの同値類にクエリ以外の楽曲が含まれないため, 楽曲が推薦できない, 2) 決定属性を選択する際にクエリの頻出部分系列数が 3 種類未満であったため, 決定属性を作成することができず, 推薦できない. 改善案として, 部分系列別の出現頻度に関する閾値を下げることで, 楽曲ごとの頻出部分系列数が増加し, よりクエリと似た部分系列を持つ楽曲を推薦できるのではないかと考えられる.

7 まとめ

本研究では, 関係性マイニングを用いた楽曲推薦手法の提案を行った. 実験結果から, 現段階では, ユーザが指定したクエリに基づき推薦を行えた場合はユーザの好みがある程度反映した楽曲が推薦されている. しかし, 推薦不能となる場合も存在するため, 今後の課題として, 推薦不能を回避するための手法の検討や, 他手法との比較実験による本手法の検証が挙げられる.

本研究の一部は科研費基盤研究 (C) (25330315) の助成を受けたものである.

参考文献

- 1) Xiaoyuan Su and Taghi M. Khoshgoftaar: A Survey of Collaborative Filtering Techniques, *Advances in Artificial Intelligence*, Vol. 2009, Article ID 421425 (2009).
- 2) 森 典彦, 田中 英夫, 井上 勝雄 編: ラフ集合と感性, 海文堂出版 (2004).
- 3) Ken Kaneiwa and Yasuo kudo: A sequential pattern mining algorithm using rough set theory, *International Journal of Approximate Reasoning*, **52**(6), 881/893 (2011).
- 4) 工藤 康生, 村井 哲也: ラフ集合による関係性マイニングの構想, 第 29 回ファジィシステムシンポジウム講演会論文集, 33/36 (2013).
- 5) 長久保大輔: 系列データを対象としたラフ集合に基づく情報推薦手法, 平成 25 年度 室蘭工業大学大学院工学研究科 情報電子工学系専攻 修士論文 (2014).
- 6) 木村光樹, 高須淳宏, 安達淳: 類似文字列検索のための Suffix Tree を用いた可変長 N-Gram, 情報処理学会第 73 回全国大会, 1-775/1-776 (2011).
- 7) Yasuo Kudo and Tetsuya Murai: A Parallel Computation Method for Heuristic Attribute Reduction Using Reduced Decision Tables, *J. Adv. Comp. Intel. Inform.*, **17**(3), 371/376 (2013).
- 8) Yasuo Kudo and Tetsuya Murai: An Evaluation Method of Relative Reducts Based on Roughness of Partitions, *Int. J. Cogn. Inform. Natur. Intel.*, **4**(2), 50/62 (2010).